

Mit MEDPILOT auf dem Weg ins Semantic Web

On the way to semantic web – with MEDPILOT

Abstract

Since 2001 MEDPILOT has developed from a usual virtual library to a powerful portal for medicine and health sciences. Today the local catalogues of the ZB MED are integrated with external databases to a unified discovery interface. This article gives an overview of the structure and features of MEDPILOT.

Besides a semantic-linguistic search functionality the integrated OPAC services are going down well with the users. This topic as well as the usage of semantic techniques on structured and unstructured data are the main content parts of the article.

Keywords: MEDPILOT, semantic search, OPAC integration, Morphosaurus, ZB MED, portal, virtual library, medicine, health

Zusammenfassung

Seit 2001 hat sich MEDPILOT von einer klassischen virtuellen Fachbibliothek hin zu einem leistungsfähigen Fachportal für die Fächer Medizin und Gesundheit weiter entwickelt. Heute werden die lokalen Kataloge der ZB MED mit externen Datenbanken zu einem Discovery-Service verknüpft und als einheitliche Rechercheplattform dem Benutzer angeboten. Über den Aufbau und die Funktionsweise von MEDPILOT wird ein Überblick gegeben.

Neben dem Anbieten einer semantisch-linguistischen Suche findet die Integration von OPAC-Funktionalitäten in MEDPILOT eine starke Nutzung. Dieses Thema wie auch die Integration semantischer Techniken auf strukturierte und nicht strukturierte Daten bilden die wesentlichen Inhaltspunkte des Beitrags.

Schlüsselwörter: MEDPILOT, semantische Suche, OPAC-Integration, Morphosaurus, ZB MED, Portal, virtuelle Fachbibliothek, Medizin, Gesundheit

1. Der Baukasten MEDPILOT

Aufbau MEDPILOT

Der Baukasten MEDPILOT (Abbildung 1) basiert in wesentlichen Teilen auf einem modularen Konzept, welches aus erprobten Komponenten und standardisierten Schnittstellen besteht. Die wesentlichen Säulen in der Software-Entwicklung von MEDPILOT sind an der ZB MED programmierte Software (vorrangig Perl), das Nutzen von Modulen und Komponenten (z.B. Suchmaschinen, CPAN) sowie klar definierte Schnittstellen (nach Möglichkeit standardisierte Datenformate) zwischen ihnen.

Dem Nutzer wird ein multifunktionales Frontend angeboten, welches als Hauptmerkmal die Recherche im Suchraum Medizin, Gesundheit als Discovery Service beinhaltet. Im Zentrum der Entwicklung von MEDPILOT steht die Gebrauchstauglichkeit (Usability) des Portals. Dafür wer-

den in regelmäßigen Abständen Usability-Studien durchgeführt, die den Fokus auf verschiedene Nutzergruppen legen: Wissenschaftler, Ärzte, Studierende, etc.

Im Rechercheportal wird durch den Nutzer eine indexbasierte Suche angestoßen. Die Anfragen werden nach einer Normalisierung über eine Schnittstelle an die Suchmaschine weiter gegeben. Die Suchmaschine trägt den Namen „Morphosaurus“ und wird von der Firma „Averbis“ in Freiburg entwickelt [1] (siehe auch „Semantische Suche in MEDPILOT“). Das Ergebnis wird durch einen an der ZB MED entwickelten Konverter aufbereitet, interpretiert und an das Portal zurück geliefert. Dort werden dann neben der übersichtlichen Darstellung der Suchergebnisse (Standardsortierung: Relevanz) die komplette Navigation erzeugt und weitere Dienste wie zum Beispiel Verfügbarkeiten zu Online-Dokumenten, die Dokumentlieferung und ganz aktuell auch die komplette Funktionalität des Kölner OPACs der ZB MED eingebunden. Die Suche findet

Christoph Poley¹

¹ Deutsche Zentralbibliothek für Medizin (ZB MED), Köln, Deutschland

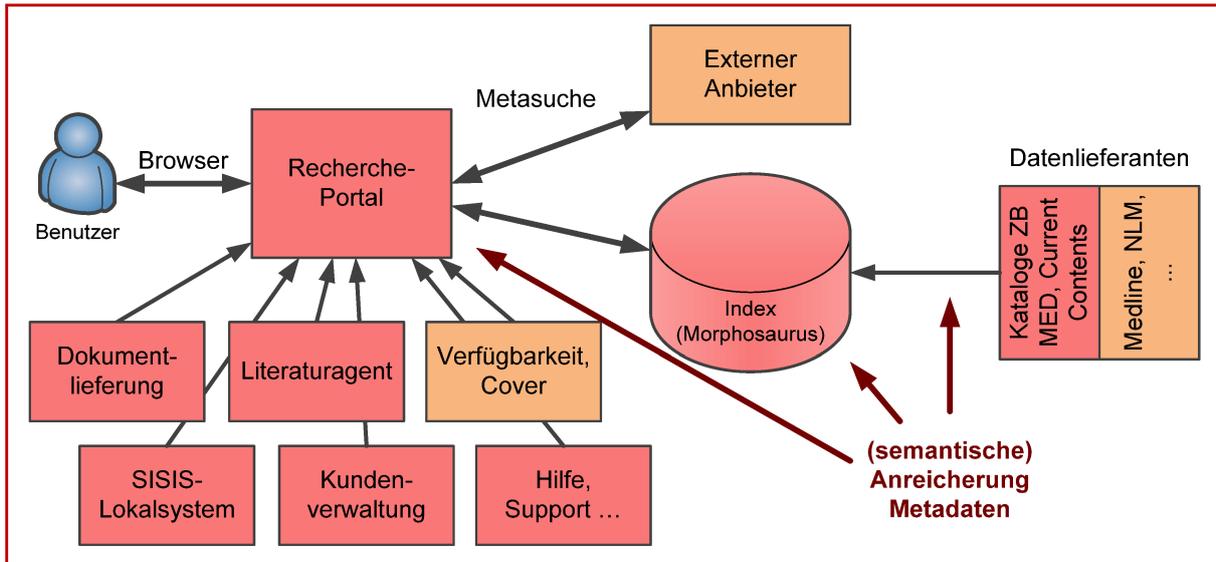


Abbildung 1: Baukasten MEDPILOT

komplett auf internen Servern der ZB MED statt. Der Datenindex ist durch ein schnelles Netzwerk mit dem Portal verbunden.

Eine wesentliche Aufgabe der ZB MED besteht darin, die Metadaten aus Datenquellen für den Index zu normalisieren und über Konverter für die Suchmaschine aufzubereiten. Darin inbegriffen sind Automatismen für das ständige Aktualisieren des Datenindexes und das Bilden von Updates (neue Datensätze, Änderungen, Löschedatensätze). Als Datenquellen dienen gleichermaßen Produkte, die durch die ZB MED erzeugt werden (Kataloge Köln und Bonn, Current Contents, German Medical Science) sowie eine Reihe von Datenbanken, die von extern bezogen werden. Hier sind vor allem Medline (PubMed), der Katalog der NLM, Dissertationen über Dissonline und Kataloge von Partnereinrichtungen zu nennen. Es werden ständig neue Datenquellen in den Index überführt. Als nächstes Ziel sollen weitere Verlagsdatenbanken aus den Bereichen Medizin und Gesundheitswesen eingebunden werden (z.B. Thieme, Karger).

Neben der indexbasierten Suche existiert parallel auch noch die klassische Metasuche, wie sie bereits seit dem Launch von MEDPILOT 2001 existiert [2]. Hier werden Suchanfragen „on the fly“ an den jeweiligen Datenbankanbieter gesendet und deren Ergebnisse wieder in das Portal überführt. Dieser Teil der Suche wird im Portal mit „Externe Datenquellen“ bezeichnet und sukzessiv durch die Überführung der Metadaten in den Datenindex abgelöst. Nur so können die Antwortzeiten der Suchanfragen möglichst gering gehalten werden. Außerdem können die Ergebnisse aus der Metasuche nicht integriert dargestellt werden, da es nicht möglich ist, zu wissen, ob ein Datensatz X aus Datenbank A eine höhere Relevanz besitzt als Datensatz Y aus Datenbank B oder womöglich Datensatz Z aus Datenbank C. Ein weiterer wichtiger Grund gegen den Einsatz der Metasuche ist die fehlende semantisch-linguistische Suche, die mit der verwendeten Suchmaschine Morphosaurus realisiert ist. Dennoch ist ein Ende der Metasuche nicht abzusehen, da aus rechtlichen oder fi-

nanziellen Gründen nicht alle Daten(-quellen) für den Index bereit gestellt werden können.

„Ich will Volltexte!“ – Dienste des Portals

Ein zentrales Anliegen jeden Nutzers ist es, möglichst schnell und einfach an die gewünschten Informationen, sowohl lizenzierte als auch kostenfreie Dokumente, zu gelangen. Oder ganz einfach umgangssprachlich formuliert: „Ich will Volltexte!“

Um zum Volltext zu gelangen, bietet MEDPILOT eine Reihe von Dienstleistungen an. Historisch gesehen zuerst ist die traditionelle Dokumentlieferung der ZB MED zu nennen. Je nach Nutzeranforderung werden Artikel oder Ausschnitte aus Zeitschriften oder Büchern eingescannt und gegen Gebühr verschickt. Bedingt durch das World Wide Web und den Anspruch, Informationen im günstigsten Falle mit einem Klick zu erhalten sowie durch die Novelle des Urheberrechtes 2008 hat die Anzahl der bestellten Dokumente in den vergangenen Jahren abgenommen.

Dennoch werden in MEDPILOT jedes Jahr mehr als eine Million Suchanfragen gestellt, Tendenz steigend. Die Lieferung von Dokumenten wird zunehmend durch elektronische Angebote ersetzt bzw. ergänzt. Neben der Einbindung von URLs freier Angebote (Dissertationen, freie Zeitschriften, Bücher, Artikel) werden Aufsätze und Zeitschriften standortunabhängig durch den Linkingdienst der Elektronischen Zeitschriftenbibliothek (EZB) verknüpft. Hier erfährt der Benutzer bereits in der Listenansicht der Trefferanzeige, ob eine entsprechende Lizenz vorliegt. Gleichzeitig werden Informationen zur Qualität der Links (Volltext, Inhaltsverzeichnis, Homepage der Zeitschrift) angezeigt. Zusätzlich bietet MEDPILOT für Bibliotheken die Möglichkeit, die Suchergebnisse mit einem lokalen Linkresolver zu verknüpfen. Mit der „Google Books Preview“ kann bei Büchern in vielen Fällen auch eine weitere Alternative im Zugang zu Covern und Volltexten angeboten werden.



Abbildung 2: Verfügbarkeiten in MEDPILOT

Weiterhin sind in MEDPILOT eine Reihe von Mechanismen integriert, die es dem Benutzer erleichtern, zu den gewünschten Daten und Informationen zu gelangen: Verknüpfungen von Datensätzen untereinander, Links in die Originaldatenbanken, durchsuchbare Inhaltsverzeichnisse sowie ein Datenexport z.B. für den Einsatz von Literaturverwaltungsprogrammen. Ein Literaturagent informiert zudem über neue Suchergebnisse nach einer gespeicherten Suche.

2. OPAC-Integration

MEDPILOT hat zum Ziel, dem Nutzer EINEN zentralen Rechercheeinstieg zu allen Angeboten der ZB MED zu bieten. Um diesem Ziel einen Schritt näher zu kommen, wurde entschieden, die OPAC-Funktionalitäten des SISIS-Sunrise-Bibliotheksmanagementsystems (BMS) vollständig in MEDPILOT zu integrieren [3]. Zwar wurden bislang bereits die Titeldaten der Kataloge in MEDPILOT angezeigt, aber für Vormerkungen oder weitere Kontofunktionen wurde eine Verlinkung in den ZB MED OPAC angeboten. Historisch bedingt, werden die Medien an der ZB MED in zwei getrennten BMS gehalten (Köln: Medizin. Gesundheit., Bonn: Ernährung. Umwelt. Agrar.), deshalb wurde dem Fächerspektrum von MEDPILOT Rechnung tragend zuerst die Vereinigung mit dem Kölner OPAC in Angriff genommen.

Ideen und Anwenderfälle

Zunächst wurde für die Realisierung der OPAC-Integration ein Aufgabenkatalog erarbeitet, als Produkt einer intensiven Anforderungsanalyse. Daran waren an der ZB MED mehrere Kolleginnen und Kollegen aus verschiedenen Fachbereichen und der Informationstechnologie beteiligt. Insgesamt wurden vier verschiedene Hauptanwenderfälle modelliert:

1. Anzeige der lokalen Verfügbarkeit des Kölner Bestandes
2. Realisierung von Interaktionen zwischen Benutzer und Lokalsystem: Vormerkung, Verlängerung, Bestellung (aus dem Magazin der ZB MED)
3. Authentifizierung der Benutzer an die SISIS-Kundendatenbank durch MEDPILOT
4. Verwaltung von SISIS-Kundendaten durch MEDPILOT

In weiteren Schritten wurden die Anwenderfälle weiter verfeinert. Die Interaktionen mit dem SISIS-OPAC dienten als Grundlage für die Feingliederung der Anwenderfälle.

Suche

Ohne ein besonderes Augenmerk darauf zu legen, wurde ein erster Teil der OPAC-Integration durch die Einbindung von Titeldaten bereits mit der Einführung von MEDPILOT realisiert. Die Anforderungsanalyse zeigte, dass wertvolle Ergänzungen für die Recherche in MEDPILOT gewünscht waren, wie neue Felder und das Beachten von Besonderheiten bei der Verarbeitung von Anfragen an die Suchmaschine. Der Datenindex von MEDPILOT wird täglich aktualisiert, das erweist sich für die Recherche in Datenfeldern als komplett ausreichend.

Verfügbarkeiten

Im Gegensatz zur Suche in den Titeldaten, müssen lokale Verfügbarkeiten (Abbildung 2) immer in Echtzeit angefragt werden.

Szenario: Benutzer A sucht nach Buch X und sieht, dass es lokal unter Signatur Z „ausleihbar“ ist. Er leiht es aus. Wenig später recherchiert ein weiterer Benutzer B nach Buch X und würde, wenn die Verfügbarkeitsdaten nur täglich aktualisiert werden würden, das Buch immer noch als „ausleihbar“ vorfinden, was aber nicht mehr den Tatsachen entspricht.

Dem Rechnung tragend, nutzt MEDPILOT einen Webservice, um diese Informationen live aus dem SISIS-System abfragen zu können. Dabei werden die einzelnen Datensätze mit weiteren Informationen angereichert, die nicht mit in den Index aufgenommen werden, beispielsweise mit Lokaldaten zu erschienenen Zeitschriftenheften (bei Zeitschriftenaufnahmen) oder Informationen über den Standort für die Ausleihe oder Bestellung. Im ersten Schritt wurden die lesenden Zugriffe auf das Lokalsystem realisiert, getestet und anschließend ab Ende 2011 in den Produktivbetrieb von MEDPILOT übernommen.

Konto- und Bestellfunktionen

Um Konto- und Bestellfunktionen des OPACs in MEDPILOT realisieren zu können, ist es notwendig, einen schreibenden Zugriff auf das BMS zu ermöglichen. Dieser wurde über einen Webservice realisiert. Einher geht damit die Verbindung der Kundendaten von MEDPILOT (für die Dokumentlieferung) mit den Nutzerdaten des BMS, um die schreibenden Aktionen aus MEDPILOT heraus auf das BMS zuzuordnen. Die Realisierung erfolgt mit der Hinterlegung wichtiger SISIS-Kundendaten auf das jeweils passende MEDPILOT-Konto, falls bereits vorhanden, und

Nr.	Titel, Verfasser	Leihfrist
1	Berliner Amnesie-Test Metzler, Peter; Voshage, Jürgen; Rösler, Petra 3576441-01 / 2012 A 3655 Vormerkung stornieren	an 1. Stelle gültig 05.10.2012 - 30.11.2019
2	Best test Ferri, Fred F. 3034478-01 / 2009 A 2083 Vormerkung stornieren	an 1. Stelle gültig 05.10.2012 - 30.11.2019
3	Klinische Tests an Knochen, Gelenken und Muskeln Buckup, Klaus	an 1. Stelle gültig 05.10.2012 - 30.11.2019

Abbildung 3: Vormerkungen im Nutzerkonto in MEDPILOT

einem asynchronem Abgleich bei der Anmeldung bei MEDPILOT (Abbildung 3).

Damit einher geht auch die Verwaltung der Kundendaten durch MEDPILOT, da dafür in naher Zukunft der SISIS-OPAC als Oberfläche nicht mehr angeboten wird.

Tests und Start

Im Anschluss der Realisierung erfolgten ausgiebige Tests, zuerst im Hause der ZB MED. Gleichzeitig fanden umfangreiche Schulungen der Mitarbeiter statt. Dieser ständige Austausch mit den Kolleginnen und Kollegen half, bis dahin unerkannte Fehler zu beheben und fehlende Anforderungen zu realisieren.

Vor der Einführung der Neuerungen wurde eine Usability-Studie mit Nutzern der ZB MED durchgeführt: Ärzte, Wissenschaftliche Mitarbeiterinnen und -mitarbeiter sowie Studierende. Als geeignete Form zeigten sich qualitative Interviews, gekoppelt mit der Anwendung der Software-suite „Morae“. Damit wurden 17 herausgearbeitete Aufgabestellungen, wie „Sie sind an einem Dokument interessiert, das bereits entliehen ist. Wie gehen Sie vor?“ mit Aufzeichnungen von Ton, Bild und Interaktion der jeweiligen Probanden in Verbindung gebracht, was Rückschlüsse auf das Verhalten im Umgang mit den integrierten OPAC-Funktionalitäten in MEDPILOT lieferte. Die Ergebnisse dieser internen Studie fließen in die laufenden Arbeiten an MEDPILOT ein.

Der Umstieg vom SISIS-OPAC auf MEDPILOT erfolgt in mehreren Schritten. Bereits Mitte 2012 wurden die neuen Funktionalitäten in MEDPILOT zur Nutzung freigegeben. Seit dem ist eine kontinuierlich zunehmende Nutzung der OPAC-Funktionalitäten in MEDPILOT zu verzeichnen, was auf eine gute Akzeptanz durch die Nutzer schließen lässt. Mit Einführung der neuen SISIS-Version 4.1 Anfang 2013 erfolgt der abschließende Umstieg auf MEDPILOT und damit verbunden die Abschaltung des klassischen SISIS-OPACs.

3. Semantische Suche in MEDPILOT

Semantische Suche

Die Semantik (Bedeutung) und damit verbunden das Semantische Web finden sich in der Literatur als Buzzwords oder Modewörter wieder. Auch als Web 3.0 bezeichnet, werden damit im Allgemeinen Internetangebote in Verbindung gebracht, die aufgrund von Wissen aus strukturierten und unstrukturierten Inhalten zusätzliche Informationen maschinell anreichern und Beziehungen zwischen ihnen zugänglich machen.

Begriffserläuterung

Semantik (Wikipedia, 25.10.2012, <http://de.wikipedia.org/wiki/Semantik>): *Semantik* (von Altgriechisch σημαίνειν *sēmaínein* „bezeichnen“), auch *Bedeutungslehre*, nennt man die Theorie oder Wissenschaft von der *Bedeutung der Zeichen*. Zeichen können in diesem Fall Wörter, Phrasen oder Symbole sein. Die Semantik beschäftigt sich typischerweise mit den Beziehungen zwischen Zeichen und Bedeutungen dieser Zeichen.

Semantik aus unstrukturierten Texten

Seit der Version 3.0 von MEDPILOT bietet die Suchmaschine die semantisch-linguistische Suche (Produkt „Morphosaurus“) der Firma Averbis (Freiburg) an. Im Gegensatz zu semantischen Beziehungen in Wissensdatenbanken und zwischen ihnen basiert diese Technologie auf dem Extrahieren von relevanten Informationen aus unstrukturierten Texten. Das Ziel dieser Verfahren ist ein Liefern vollständiger Rechercheergebnisse. Werden bei herkömmlichen Verfahren die Suchanfragen nahezu un bearbeitet mit den Daten aus den Suchmaschinen abgeglichen, so geht deren Aufbereitung beim Einsatz des Morphosaurus weiter.

So können neben exakten Treffern auch Ergebnisse aus anderen Sprachen, Experten-/Laiensprache, Komposita, Derivaten, etc. gefunden werden. Dieses Ziel wird durch eine linguistische Zerlegung der Wörter in den Suchanfra-

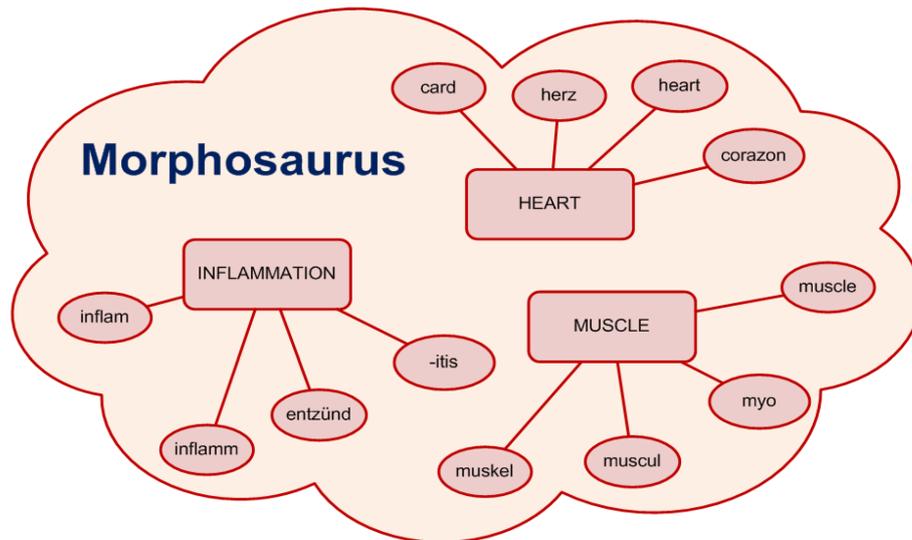


Abbildung 4: Beispiel Herzmuskelentzündung

gen in ihre Wortbestandteile (Morpheme) erreicht, die dem MeSH – **M**edical **S**ubject **H**eadings – entnommen sind. Durch einen Thesaurus kann diese überschaubare Anzahl der Morpheme auf sprachunabhängige Konzepte abgebildet werden. Diese linguistisch generierten Informationen werden zusätzlich in den Datenindex aufgenommen. Der gleiche Mechanismus greift auch bei der Aufbereitung der Suchanfragen, so dass nun ein Matching auf die sprachunabhängigen Konzepte erfolgen kann und damit Ergebnisse geliefert werden, die zwar nicht den exakt gesuchten Begriff enthalten, aber z.B. dessen Übersetzung.

Beispiel: Herzmuskelentzündung

Das in Abbildung 4 dargestellte Wort „Herzmuskelentzündung“ wird in die Morpheme „Herz“, „Muskel“ und „Entzündung“ zerlegt und auf die Wortstämme „HEART“, „MUSCLE“ und „INFLAMMATION“ abgebildet. So wird neben Übersetzungen wie „inflammation of the heart muscle“ auch der Fachterminus „Myokarditis“ gefunden.

Semantik aus strukturierten Texten

Im Sinne der semantischen Anreicherung von Daten können vereinfacht an drei Stellen im „Baukasten MEDPILOT“ zusätzliche Informationen eingebracht werden.

1. Konvertierung der Quelldaten in das Format für die Suchmaschine zur Indexierung: Bereits hier können durch Meshups und einfacher Textanalyse zusätzliche Identifikatoren in die Datensätze aufgenommen werden. Der Aufbau der Wissensbasis dafür kann unabhängig asynchron zur Konvertierung der Daten erfolgen und zum Beispiel in Schattendatenbanken münden, die für die Weiterverarbeitung der Daten herangezogen werden.
2. Indexierung durch den Morphosaurus. Hauptaugenmerk wird derzeit hier auf die Analyse von Texten gelegt.

3. Anreicherungen durch das Portal: Nicht alle semantischen Informationen müssen in den Index mit einbezogen werden, da sie auch ergänzende Inhalte bieten (z.B. URLs), die für die Suche nicht relevant sind.

Abbildung 5 zeigt einen Ausschnitt aus der Welt des Semantic Webs, die LOD-Cloud. Es wird deutlich, dass einzelne Wissensdatenbanken (als Kreise dargestellt) nicht für sich alleine dastehen, sondern dass es zwischen ihnen Beziehungen aufgrund von Bedeutungen gibt.

Semantik – ein gerichteter Graph

Zusammenhänge innerhalb von einer Wissensdatenbank können als gerichteter Graph aufgefasst werden. Damit einher geht, dass es zwischen ihnen als Knoten gerichtete Kanten gibt. Sie stehen sinnbildlich für die Beziehungen zwischen den Knoten.

Solcherlei kleinste semantische Zusammenhänge lassen sich auch als Tripel darstellen: Subjekt, Objekt, Prädikat. Als Subjekt wird der Knoten gesehen, von dem aus eine Beziehung zu einem Objekt – auch ein Knoten – ausgeht. Die Verbindung – eine gerichtete Kante – und deren Bedeutung werden als Prädikat bezeichnet.

Beispiel: In Abbildung 6 ist ein einfacher Sachverhalt aus einer fiktiven Wissensdatenbank dargestellt, wo drei Objekte in Beziehung gebracht werden. Solcherlei Beziehungen können auch maschinell lesbar dargestellt werden. Als weit verbreitete Sprache dient hier das Resource Description Framework (RDF). Oben genannter Sachverhalt könnte damit wie folgt aussehen (in RDF – Turtle-Schreibweise) und auf einem Server in einem Triple Store zur Verfügung gestellt werden:

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
<http://localhost:5000/Alfred> a foaf: Person;
  foaf:name "Alfred" ;
  foaf:knows <http://localhost:5000/Elke>;
  foaf:knows ... .
<http://localhost:5000/Elke> a foaf: Person;
  foaf:name "Elke" ;
```

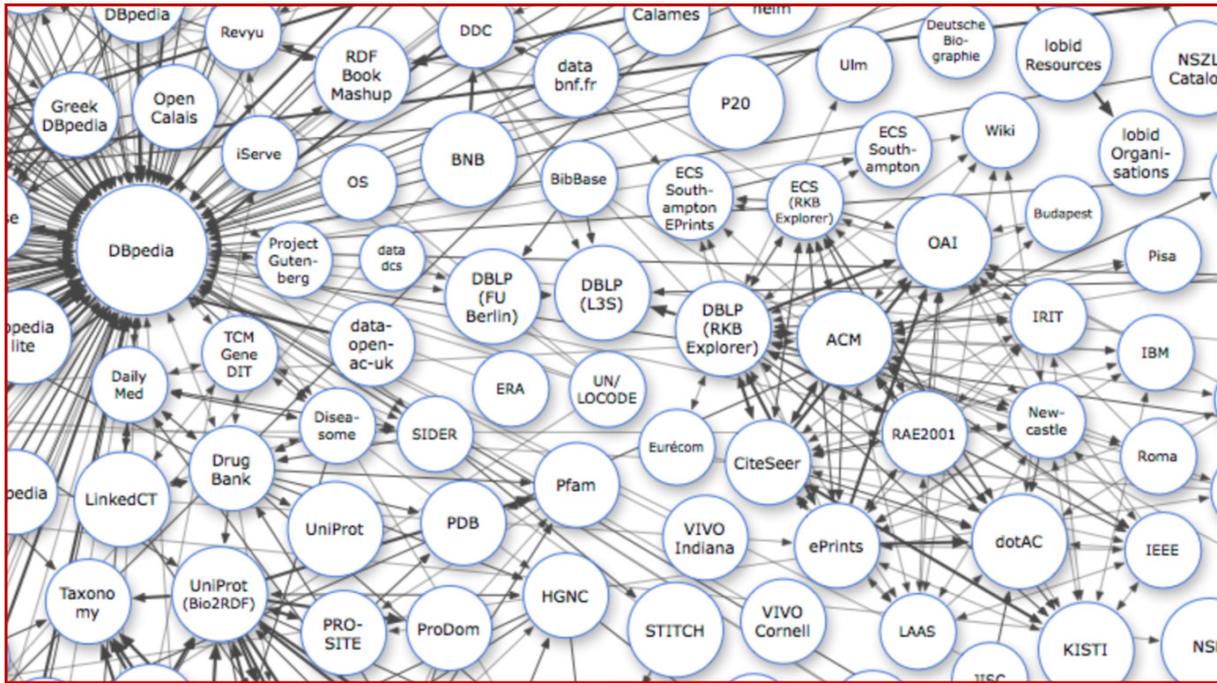


Abbildung 5: Ausschnitt aus Linked-Open-Data-Cloud (LOD-Cloud) nach <http://richard.cyganiak.de/2007/10/lod/imagemap.html> (25.10.2012)

foaf:current Project <<http://localhost:5000/Koeln>> .
 <<http://localhost:5000/Koeln>> a foaf: Document;
 foaf:name "Domstadt Köln" .

Weiterentwicklungen in MEDPILOT – Beispiel Drugbank

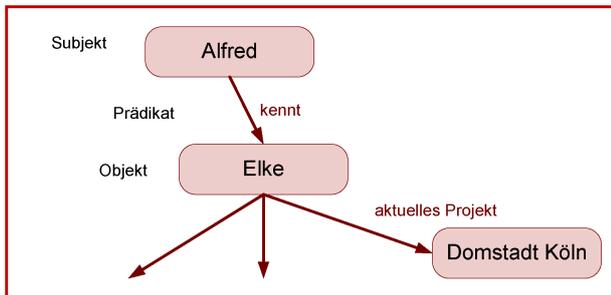


Abbildung 6: Semantik, natürliche Sprache

Als Abfragesprache hat sich SPARQL (SPARQL Protocol And RDF Query Language) etabliert.

Aktivitäten ZB MED

Seit 2010 werden von der ZB MED die Katalogdaten bereits unter <http://www.lobid.org/about> zur weiteren Verarbeitung angeboten. Die Daten liegen in RDF vor, ein Browsing ist auch möglich; Beispiel: <http://lobid.org/resource/ZDB200772-1>.

Zusätzlich werden die Current-Content-Dienste der ZB MED CC MED (für Medizin, Gesundheit) und CC GREEN (für Ernährung, Umwelt, Agrar) als SIS-Dump über das Portal CKAN angeboten (<http://thedatahub.org/dataset>).

Semantische Verknüpfungen zwischen Objekten sollen auch in Zukunft in die Entwicklung von MEDPILOT einfließen. Einige Verfahren, wie das Anreichern von Medline-Datensätzen, werden in der ZB MED schon seit Jahren praktiziert, diese sollen durch weitere sinnvolle und nützliche Informationen ergänzt werden.

Als Beispiel für den Einstieg in semantische Verknüpfungen soll die bio2rdf-Wissensdatenbank dienen. Dort könnten etwa sämtliche Knoten von Substanzen (DBxxxxxx) extrahiert und weiter verarbeitet werden. So existiert vom Knoten <http://bio2rdf.org/drugbank:DB01050> eine Verbindung (http://bio2rdf.org/drugbank_vocabulary:xref) zu einer CAS (Chemical Abstracts Service)-Nummer (<http://bio2rdf.org/page/cas:15687-27-1>). Weiterhin ist eine Verbindung in die verwandte Drugbank (Abbildung 7) existent, von wo aus weitere in Relation zum Ausgangsknoten stehende Inhalte gefunden werden können. Besonders relevant ist hier die Information, dass es sich bei der gesuchten Substanz um Ibuprofen handelt. Auch werden unter anderen Verknüpfungen in die englischsprachige Wikipedia und zu einer grafischen Aufbereitung der Substanz (<http://www.drugbank.ca/drugs/DB01050>) angeboten.

Für den praktischen Einsatz in MEDPILOT können jetzt diese Informationen bei der Indexierung von Datenquellen, z.B. Medline, herangezogen werden. Tritt beispielsweise im Titel oder im Abstract der Begriff „Ibuprofen“ oder bei den chemischen Substanzen eine CAS-Nummer auf, dann können die eben gewonnenen Informationen für die Anreicherung der Datensätze herangezogen und als Identifikatoren in den Index eingefügt werden. Die so

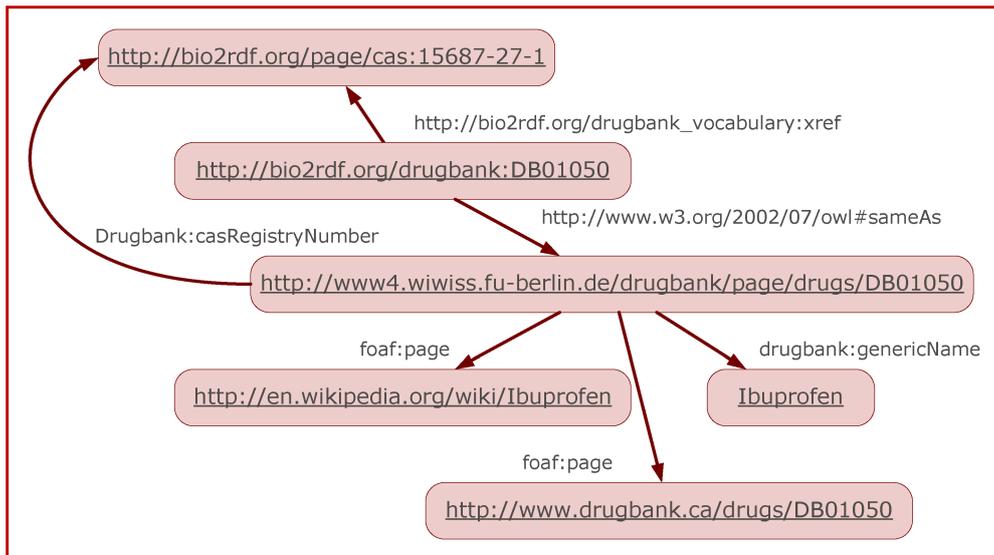


Abbildung 7: Beispiel Drugbank

```

#FORMAT: BEACON
#PREFIX: http://d-nb.info/gnd/
#VERSION: 0.1
#TARGET: http://toolserver.org/~authoritycontrol/redirect/gnd/de/{ID}
#FEED: http://toolserver.org/~authoritycontrol/beacon/dewiki.txt
#CONTACT: Christian Thiele <apper@apper.de>
#INSTITUTION: Deutschsprachige Wikipedia
#DESCRIPTION: Wikipedia-Personenartikel
#TIMESTAMP: 2012-10-26T06:18:02
4523163-1
119317079
120882965
118551647
118972332
4140847-0
119070243
4141899-2
  
```

Abbildung 8: Beacon

angereicherten Suchergebnisse lassen sich z.B. im Portal in Templates weiter aufbereiten.

GND, Beacon, VIAF

Neben dem klassischen Nutzen von Wissensdatenbanken existiert eine Reihe von Mechanismen, wie Datensätze weiter semantisch angereichert und verknüpft werden können. Dabei ist die Nutzung von Dateien im Beacon-Format (Abbildung 8) hervorzuheben. Sie helfen bei der Beantwortung der Frage, welches Rechercheangebot (Kataloge, Wikipedia, etc.) zu einem gegebenen Autor wie viele Dokumente besitzt.

Dementsprechend einfach sind die Beacon-Dateien für diesen Fall auch aufgebaut. Im Wesentlichen bestehen sie aus einer Base-URL (#TARGET) des Anbieters und den GND (Gemeinsame Normdatei)-IDs, für die es Treffer gibt. Somit lassen sich Datensätze, die eine GND-ID beinhalten, leicht verknüpfen. Dem Benutzer können so weitere Recherchemöglichkeiten angeboten werden. Die bereits

existierenden Beacon-Angebote sind frei verfügbar und werden in der Wikipedia unter <http://de.wikipedia.org/wiki/Wikipedia:PND/BEACON> angeboten.

Eine weitere sehr interessante Thematik ist die eindeutige Identifikation von Autoren. So existiert der Anwendungsfall, dass zu einem Datensatz aus einem gegebenen Suchergebnis sämtliche in MEDPILOT bekannten Datensätze des Autors gefunden werden sollen. Da die intuitive Suche über den Autorennamen nicht ausreicht (z.B. Suche nach Hans Müller), existieren verschiedene Identifikatoren für Autoren, z.B. GND, die in der Wissensdatenbank VIAF (Virtual International Authority File) zusammengefasst sind. Wenn nun ein Großteil der Datensätze aus MEDPILOT über die VIAF eindeutige Autoren-Identifikatoren erhalten, wird es möglich sein, Facetten nach Autoren aus der Ergebnisliste für ein Filtern der Suchergebnisse anbieten zu können.

4. Fazit

Nicht nur in der Informationstechnologie werden in regelmäßigen Abständen Buzzwords hervorgebracht, die in aller Munde sind, aber bei näherer Betrachtung eine Vielzahl von unterschiedlichen Auffassungen darüber vereinen. Genauso verhält es sich bei dem Thema Web 3.0 oder „Semantisches Web“. Dennoch lohnt es, sich mit dieser Thematik näher auseinander zu setzen, denn es existieren hier vor allem reale Anwenderfälle, die durch Ausnutzung der dahinter stehenden Technologien in die Praxis umgesetzt werden können.

Das Produkt MEDPILOT der ZB MED bedient sich seit geraumer Zeit verschiedener Techniken zu dem Thema, um den Benutzern einen Mehrwert an Funktionen und Inhalten zu vermitteln. Deshalb wurden die OPAC-Funktionalitäten konsequent in MEDPILOT integriert. Sie erlauben den Benutzern das Ausführen weitreichender Funktionen bei Datensätzen aus dem Katalog der ZB MED. Darüber hinaus werden diese Datensätze gleichwertig neben allen weiteren (Medline, Dissonline, CCMED, ...) dem Benutzer angeboten.

In Zukunft werden semantische Verfahren eine wichtige und verstärkte Rolle bei der Weiterentwicklung von MEDPILOT spielen, um möglichst unabhängig von Herkunft eines jeden Datensatzes die vorhandenen Informationen mit neuen Datenquellen zu verknüpfen und damit anzureichern. Für den weiteren erfolgreichen Fortbestand einer virtuellen Fachbibliothek reicht es nicht aus, Informationen aus verschiedenen Datenquellen anzubieten. Sie müssen in diesem Fall, wenn sinnvoll, intelligent durch die Ausnutzung von Semantiken in Verbindung gebracht werden. Die Nutzer werden es uns danken.

Anmerkung

Interessenkonflikte

Der Autor erklärt, dass er keine Interessenkonflikte in Zusammenhang mit diesem Artikel hat.

Literatur

1. Dzeyk W. Medizinische Literaturrecherche mit MEDPILOT, PubMed und Google. Leistungstest und Usability-Vergleich. Köln: Verlag Deutsche Zentralbibliothek für Medizin; 2011. DOI: 10.4126/38m-000000379
2. Schneider S. Das Motto lautet – Recherche und Literaturbestellung leicht gemacht! *medizin – bibliothek – information*. 2004;4(1):33-5. Verfügbar unter: http://www.agmb.de/mbi/2004_1/schneider33-35.pdf
3. Hinnenthal K, Pössel J, Schuh D. Die Deutsche Zentralbibliothek für Medizin hat ihr Suchportal für Medizin und Gesundheitswesen um einen entscheidenden Schritt weiterentwickelt: Integration der Bibliothekskontofunktion in MEDPILOT. *GMS Med Bibl Inf*. 2012;12(1-2):Doc13. DOI: 10.3205/mbi000249

Korrespondenzadresse:

Christoph Poley
Deutsche Zentralbibliothek für Medizin (ZB MED), Gleueler Str. 60, 50931 Köln, Deutschland
christoph.poley@zbmed.de

Bitte zitieren als

Poley C. Mit MEDPILOT auf dem Weg ins Semantic Web. *GMS Med Bibl Inf*. 2012;12(3):Doc22.

DOI: 10.3205/mbi000258, URN: urn:nbn:de:0183-mbi0002581

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/mbi/2012-12/mbi000258.shtml>

Veröffentlicht: 20.12.2012

Copyright

©2012 Poley. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.