RESIST – Regensburger Signalpfad Informationssystem[1]

Prof. Dr. Rainer Hammwöhner
Lehrstuhl für
Informationswissenschaft
Universität Regensburg
93040 Regensburg

Prof. Dr. Rainer H. Straub
Forschungsleiter
Klinik und Poliklinik für Innere Medizin I
Universität Regensburg
93040 Regensburg

1. Einleitung

In diesem Beitrag soll ein im Aufbau befindliches Forschungsvorhaben beschrieben werden, das in Kooperation von Informationswissenschaftlern und Klinikern der Universität Regensburg durchgeführt wird. Ziel dieses Projekts ist die Entwicklung eines Signalpfadinformationssystems neuen Typs, das einerseits einen erweiterten Kreis an Phänomenen erfassen kann und anderseits Instrumente zur Wissenschaftskommunikation zur Verfügung stellt.

Das Auffinden von Signalpfaden[2] ist ein aktives Forschungsgebiet in Biologie, Biochemie und Medizin. Signalpfaddatenbanken dokumentieren die Ergebnisse dieser Forschung. Der Schwerpunkt wird derzeit auf die Erforschung und Dokumentation intrazellulärer Prozesse gelegt. Prozesse auf extrazellulärer und organverbindender Ebene bleiben weitgehend unberücksichtigt, so dass ein Brückenschlag zwischen biochemisch-molekularbiologischer und klinischer Forschung unterbleibt. Wichtige Forschungsergebnisse etwa aus Immunologie oder Endokrinologie können nicht einfach mit den Datenbankinhalten in einen Zusammenhang gebracht werden.

An dieser Stelle setzt das hier skizzierte Forschungsvorhaben an. Zunächst werden Instrumente für die integrierte Beschreibung der aus Zellforschung, Immunologie, Endokrinologie usw. erwachsenden Forschungsergebnisse geschaffen. Dazu sind vorhandene Ontologien zu sichten, zu erweitern und zu integrieren. Beschreibungsformate für intrazelluläre, interzelluläre und gesamtorganismische Phänomene werden entwickelt. Da die Bedeutung von Signalmolekülen je nach Zuordnung zu einem Kompartiment (Körper-, Organ-, oder Zellsegment) wechseln kann, wird eine Erfassung dieser Strukturen erforderlich. Um den im interdisziplinären Kontext jeweils wechselnden Anforderungen an den Abstraktionsgrad der Beschreibung und die Detaillierung der Modellierung gerecht zu werden, wird die Möglichkeit einer mehrschichtigen Annotation der Daten vorgesehen. So können auch Beschreibungslücken oder Inkongruenzen in der Forschung zwischen den Teildisziplinen abgebildet werden. Mechanismen zur strukturorientierten Recherche in den Datenbeständen und zur Visualisierung der Ergebnisse sind vorgesehen.

Insbesondere in dynamischen Wissenschaftsgebieten stellt der lange Publikationsweg ein hohes Hemmnis für die effiziente Wissenschaftskommunikation dar. Um hier einen Ausweg

---

[1] Dieser Text ist erschienen in: Knowledge eXtended. Die Kooperation von Wissenschaftlern, Bibliothekaren und IT-Spezialisten. Forschungszentrum Jülich. Proc. 3. Konferenz der Zentralbibliothek. Schriften des Forschungszentrums Jülich, 2005, S. 237-249.

[2] Signal Transduction Pathways, Metabolic Pathways, Biochemical Pathways

zu schaffen, wird RESIST mit einer offenen Kommunikationsplattform versehen. Hier können Wissenschaftler Forschungshypothesen und Experimentaldesigns zur Debatte stellen. Es ist wichtig, dass in der Debatte eine exakte Bezugnahme auf die repräsentierten Strukturen ermöglicht wird. Als Folge dieses offenen Charakters wird das Informationssystem Fakten und Hypothesen unterschiedlicher Validität und Reichweite enthalten. Diesem Umstand wird durch die Zuweisung differenzierter Qualitätsmerkmale Rechnung getragen.

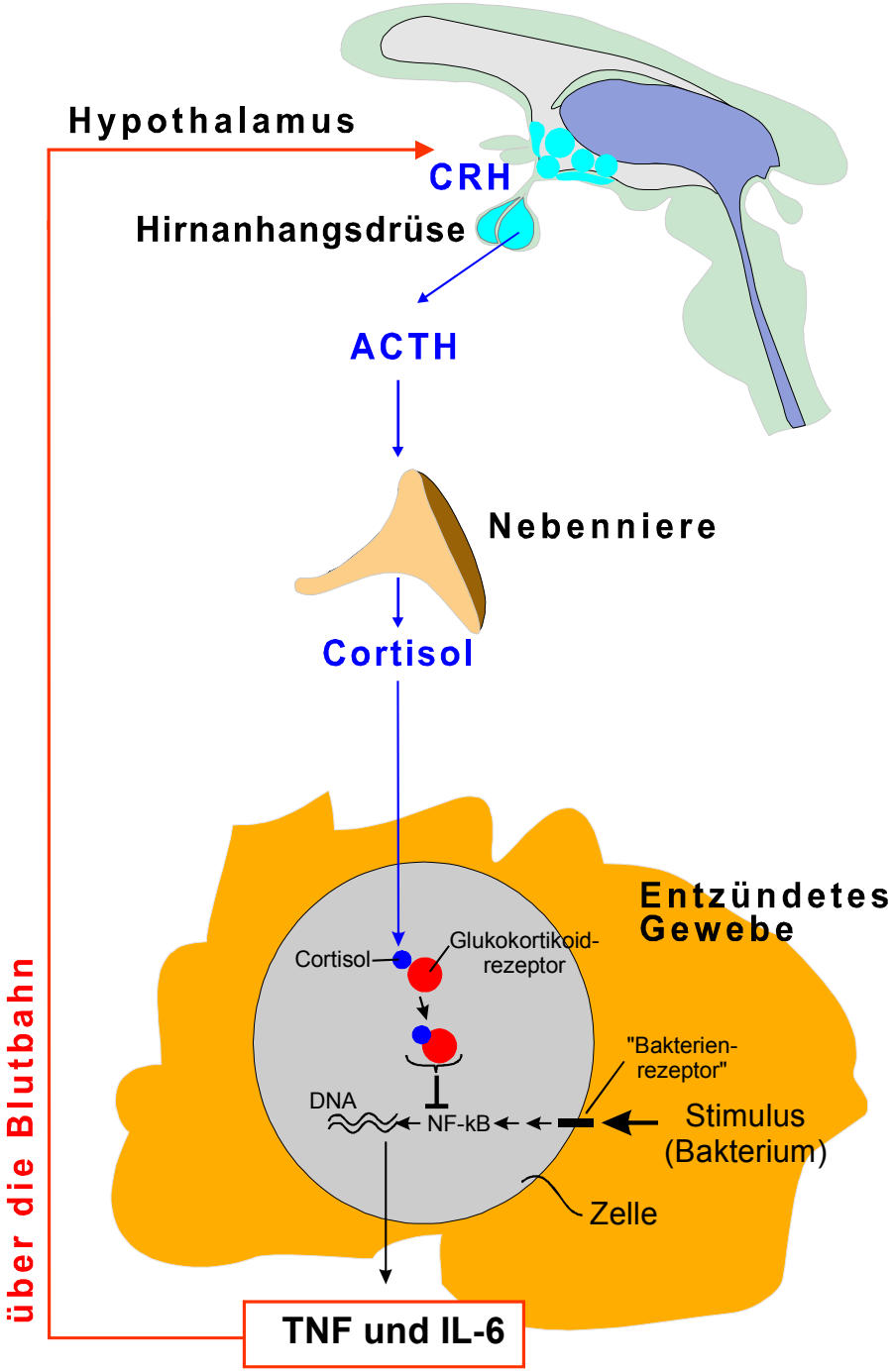

Abbildung:

Verbindung eines gesamtorganismischen und eines intrazellulären Signalpfades.

Ein Bakterium stimuliert über einen „Bakterienrezeptor“ eine Zelle in einem entzündeten Gewebe. Dabei wird der intrazelluläre Faktor NF-kB produziert, der wiederum die Produktion von Tumor-Nekrose-Faktor (TNF) und Interleukin-6 (IL-6) bewirkt. Diese Faktoren gelangen aus dem entzündeten Gewebe in den gesamten Organismus (Blutbahn) und stimulieren den Hypothalamus, eine Region innerhalb des Gehirns. Dort stimuliert TNF das Corticotropin-Releasing-Hormon (CRH), das wiederum das Hormon der Hirnanhangsdrüse ACTH stimuliert. ACTH stimuliert in der Nebenniere Cortisol, das wiederum über die Blutbahn in das entzündete Gewebe gelangt. Dort hemmt Cortisol nach der Bindung an den Glukokortikoidrezeptor die Wirkung des NF-kB.

2. Stand der Wissenschaft und Technik

Im folgenden soll ein kompakter Überblick über die aktuelle Entwicklung in den relevanten Forschungsgebieten vermittelt werden.

2.1 Forschung im Bereich der Signalpfade

Hinsichtlich der zielgruppenspezifischen Erstellung, Aufbereitung und Verbreitung von Informationsinhalten konzentrieren sich die Bemühungen zur Zeit besonders auf intrazelluläre Signalpfade, wie sie für Molekularbiologen und Biochemiker nützlich sind (Signalkaskaden, Signaltransduktionspfade, Reaktionszyklen). Derartige Signalpfad-Datenbanken sind auf der Homepage „http://home.comcast.net/~natgoodman/Pathway_Web_Sites.htm“ einsehbar. Als Beispiel für ein Signalpfad-Datenbanken hoher Qualität sei hier die Datenbank des *Kyoto Encyclopedia of Genes and Genomes* (KEGG) genannt. Trotz des unbestreitbaren großen Informationsangebotes derartiger Datenbanken bestehen dennoch Nachteile:

1. In den meisten Fällen sind medizinische Themen, die den gesamten Organismus im Blickpunkt haben, von den Betrachtungen ausgeschlossen. Die Kluft zwischen Biochemie/Molekularbiologie (*Genomics* und *Proteomics*) einerseits und klinischen Fragestellungen andererseits wird aufgrund der unterschiedlichen Abstraktionsebenen größer, obwohl genau das Gegenteil wünschenswert wäre. Zwischen Krankheitssymptomen und zellulären Vorgängen wird die stringente Verknüpfung vermisst.
2. In allen uns bekannten Fällen beschreiben die vorhandenen Signalpfad-Datenbanken intrazelluläre Vorgänge ohne Berücksichtigung von Kompartimenten[3]. Man betrachtet die Signalpfade dahingehend, als ob sie in einem einzigen Kompartiment vorhanden wären. Damit wird nicht erfasst, das ein Signalmolekül in verschiedenen Kompartimenten unterschiedlichen Wirkung haben kann.
3. Die Datenbanken sind nicht für die Nutzer offen, d.h. Nutzer können keine Eingaben vornehmen, um eigene Forschungsergebnisse zur Debatte zu stellen.

Diese Nachteile machen die bisherigen Signalpfad-Datenbanken zu einer Informationsquelle für Biochemiker, Molekularbiologen, Genomiker und Proteomiker, nicht aber für gesamtorganismisch orientierte Pharmakologen, Physiologen, Pathophysiologen, klinisch tätige Ärzte und Studenten der Biologie/Humanmedizin. Die Themen sind zu speziell und zellorientiert. Die Verknüpfungen zwischen gesamtorganismischen Symptomen (z.B. Kopfschmerz) und zellbasierter Forschung sind nicht abgebildet. Hier soll das anvisierte Projekt mit Hilfe einer neuen Form der Signalpfad-Datenbank in einem Internet-basierten, offenen, digitalen Informationssystem Abhilfe schaffen. Hierbei ist es ein Hauptinteresse der Antragsteller, die vorhandenen intrazellulär orientierten Signalpfad-Datenbanken mit dem neuen System zu verknüpfen. RESIST kann insofern als integrierende Metaebene über existierenden Datenbanken betrachtet werden, aus welchen via RESIST Fakten abgerufen werden können. Insofern ersetzt RESIST keinesfalls die bekannten Datenbanken, sondern integriert das neue offene Konzept mit den bekannten Faktendatenbanken. Bei der Entwicklung von RESIST sind verfügbare Open-Source-Projekte im Bereich der Signalpfad-Datenbanken – etwa die Reactome-Datenbank (Joshi-Tope et al. 05) – zu berücksichtigen. So wird einerseits der Entwicklungsaufwand reduziert, die Kompatibilität zwischen bestehenden Ansätzen andererseits erhöht.

[3] Je nach Betrachtungsebene umgrenzte Räume einer Zelle, eines Organs oder des Organismus.

2.2 Dokumentation von Signalpfaden

Als Grundlage für die Dokumentation der Signalpfade werden Ontologien – im Sinne der Spezifikationen des Semantic Web – angesehen, wie sie von Forschergruppen und Fachverbänden vorgeschlagen werden (vgl. auch Lewis 2005). Vorschläge für Ontologien in Biologie, Bioinformatik und Medizin liegen vor[4]. Es sind bereits mehrere Ontologien verfügbar[5], die den relevanten Gegenstandsbereich zumindest partiell abdecken. Diese sind gegebenenfalls zu erweitern und an den Bedarf des konzipierten Informationssystems anzupassen[6]. Insbesondere sind funktionelle Aspekte zu berücksichtigen - vgl. (Takai-Igarashi, Mizoguchi 2003) oder (Smith et al. 2005 a) – aber auch einfache anatomische Zusammenhänge müssen zur Repräsentation der Kompartimente berücksichtigt werden (Smith et al. 2005b). Eine Basis für diese Erweiterung kann durch eine Einbettung in die Unified Medical Language geschaffen werden. Die Möglichkeit der Abbildung der Gene Ontology auf UMLS wurde bereits demonstriert (Lomax, McCray 04). Einen Überblick über verschiedene Datenmodelle für die Repräsentation von Biochemical Pathways vermitteln (Deville et al. 03). Ein objektorientiertes Datenmodell für eine objektorientierte Datenbank stellt (Schacherer 01) vor, und gibt damit wertvolle Hinweise für die Realisierung derartiger Systeme.

Auch für die Erfassung von Qualitätsinformationen über die im System verwalteten Fakten – wissenschaftliche Evidenz usw. – muss eine Ontologie bereitgestellt werden. Hier ist auf den von (Karp 04) publizierten Ansatz zu verweisen.

Auf der Basis der zur Verfügung stehenden Ontologien können dann Signalpfade beschrieben, bzw. Publikationen über Signalpfade inhaltlich deskribiert werden. Dazu werden heute bereits Annotationsschemata vorgeschlagen. Hinsichtlich der Beschreibung von Signalpfaden auf unterschiedlichen Ebenen biologischer oder medizinischer Phänomene (intrazellulär, interzellulär, organbezogen, gesamtorganismisch) sind Verfahren der Mehrebenenannotation zu berücksichtigen. Manche der bereits für die Biologie vorgeschlagenen Annotationsschemata schließen Mehrebenenannotationen ein, wobei hier zumeist Abstraktionsebenen unterschieden werden (Battistella et al 2004, Paek et al 2004). Darüber hinaus sind in der Sprachwissenschaft, insbesondere in der Teildisziplin Texttechnologie entwickelte Annotationsverfahren (s. 2.5) zu berücksichtigen.

Hohe Bedeutung kommt bei diesen Aktivitäten der Datenerfassung zu. Damit der aktuelle Forschungsstand angemessen wiedergegeben werden kann, muss eine große Anzahl von Altpublikationen erschlossen, sowie der umfängliche Strom der Neupublikationen mit hoher Vollständigkeit abgedeckt werden. Eine Lösung für das Problem des Erschließungsaufwands wird vielfach in einer semi-automatischen Analyse der Texte gesehen, der eine automatische Faktenextraktion zu Grunde liegt (etwa Koike et al. 04; Karopka et al 04). Die eingesetzten Verfahren sind zumeist in der statistischen Sprachverarbeitung oder in der Erkennung von Satzmustern begründet. Diese Verfahren könnten auch zur Neuindexierung von Altpublikationen herangezogen werden. Es ist jedoch zweifelhaft, ob diese Ansätze zu ausreichend verlässlichen Ergebnissen führen.

---

[4] Zusammengetragen etwa unter dem Titel *Open Biological Ontologies* (http://obo.sourceforge.net/, zitiert 28.1.05),

[5] Etwa aus dem Gene Ontology Project (http://www.geneontology.org/, zitiert am 28.1.05), zu berücksichtigen ist aber auch die MeSH-Terminologie, das kontrollierte Vokabular von PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=mesh, zitiert am 28.1.05).

[6] Einen Überblick über die Entwicklung und Fusion von Ontologien geben (Ding, Foo 02a und 02b).

Gegenstand der biochemischen Forschung sind, wie bereits oben skizziert, zumeist intrazelluläre oder zumindest vergleichsweise lokale Prozesse. Diese Ausrichtung der Forschung spiegelt sich in der Struktur und Ausarbeitungstiefe der verwendeten Ontologien wider. Dem Kliniker wird so der Zugang zu Forschungsergebnissen, die auch für ihn relevant sein können, erschwert. Hier ist eine Brücke zwischen den beobachteten Phänomenen und verwendeten Sprachen zu finden.

Die derzeit verfügbaren Signalpfaddatenbanken werden primär durch Erschließung von begutachteten Fachpublikationen aufgebaut. Damit ist das Problem der Validität der aufgenommenen Daten – sieht man von möglichen Erschließungsfehlern ab – auf die Fachgutachter der Publikationsorgane delegiert. Als Folge übertragen sich die Probleme der traditionellen Wissenschaftskommunikation – lange Vorlaufzeiten, z.T. schwer nachvollziehbare Begutachtungsprozesse – auf dieses vom Potential her schnellere und demokratischere Medium.

2.3 Abfragesprachen für Signalpfade

Für die komplexen Graphstrukturen, die als Repräsentationen von Signalpfaden entstehen werden, muss eine adäquate Möglichkeit der Informationssuche geschaffen werden. Während zahlreiche Datenbanken bisher allein konventionelle Methoden der Abfrage – via SQL o.Ä. – anbieten, sind derzeit mächtigere Abfragesprachen in Entwicklung und Erprobung, die auf Verfahren des Graphenabgleichs beruhen (Sohler, Zimmer 2005; Pinter et al. 2005). Diese können auf allgemeine Methoden der Graphensuche zurückgreifen, für die bereits etablierte Werkzeuge zur Verfügung stehen (Giogno, Shasha 2002).

2.4 Visualisierung von Signalpfaden

Der graphischen Darstellung von Signalpfaden kommt eine hohe Bedeutung für das Verständnis dieser komplexen Strukturen zu. Ohne ein quasi-räumliches Modell fällt es schwer, hier einen Überblick zu wahren. Ein Pilotprojekt stellte die Visualisierung der Boehringer Poster dar, eine umfassende Darstellung biochemischer Reaktionszyklen, für die im Rahmen eines vom BMBF geförderten Projekts eine CD-basierte Fassung erstellt wurde (Kanne et al. 99). Die entwickelten Verfahren sind heute im System BioPath verfügbar (Schreiber 02). Mittlerweile ist eine Vielzahl von Signalpfaddatenbanken mit Visualisierungswerkzeugen ausgerüstet, weitere Visualisierungstools sind unabhängig von einer Datenbank verfügbar. Sie beruhen – wenn nicht ohnehin von einer intellektuellen Aufbereitung ausgegangen wird – auf graphbasierten Layoutverfahren – vgl. (Becker, Rojas 2001), (Goesmann et al. 2002), (Krieger et al. 2004).

2.5 Web Services

Es ist offensichtlich, dass das geplante Projekt zahlreiche Anknüpfungspunkte zu anderen Aktivitäten innerhalb der Erforschung biologischer Signalpfade aufweist. Wie in der obigen Abbildung schon angedeutet, sollen auch weitere Informationsdienstleistungen aus dem Web mit den innerhalb von RESIST verwalteten Daten genutzt werden können. Zu berücksichtigen sind netzbasierte Dienstleistungen, die über gut definierte Schnittstellen als Webservices angeboten werden. Zu denken ist an Dokumentenzugang (Ghandeharizadeh et al. 02), automatische Texterschließung (Vargas-Vera et al. 02), Auswertung biologischer Modelle (Rahman et al. 04), Visualisierungen (Rojdestvenski 03) etc.

Hier ist die Systemarchitektur[7] so anzulegen, dass bestehende Dienstleistungen gut integriert werden können. Die Anforderungen an die Schnittstellen müssen berücksichtigt werden, ebenso wie die eventuell erforderliche Anpassung der Inhaltsrepräsentationen.

2.6 Texttechnologische Ansätze

Aus der sich abzeichnenden Notwendigkeit heraus, die Daten auf mehreren Abstraktionsstufen zu beschreiben – etwa biochemische und gesamtorganismisch physiologische Ebene oder quantitative (vgl. Sivakumaran 2003) vs. qualitative Beschreibung – erhebt sich die Frage nach der angemessenen Gestaltung und gegenseitigen Bezugnahme der entsprechenden Deskriptoren. Eine ähnliche Problematik stellt sich in der Linguistik, wenn es um die Beschreibung von Sprach- und Textdaten geht. Auch hier sind mehrere Ebenen der Beschreibung erforderlich (vgl. etwa Sasaki 2004) – etwa eine phonologisch/graphematische Ebene, Syntax, Koreferenz, Semantik und Pragmatik (vgl Teich, Hansen 01) –, die auch nicht unverbunden nebeneinander stehen dürfen. Hier werden Verfahren der Mehrebenenannotation entwickelt, wobei jede Ebene über ihre Ontologien verfügt, die über Korrespondenzregeln verknüpft werden können. Unterstützung durch Annotationswerkzeuge wird bereitgestellt (z.B. Müller, Strube 2003). Verfahren der Mehrebenenannotation werden aber auch im Image Retrieval genutzt (Fan et al. 2004). Hier ist vor allem die Kombination der Oberflächeneigenschaften des Bildes mit Information über die Bildinhalte (Szenen, Konzepte) von Bedeutung.

Eine weiteres wichtiges Anwendungsgebiet texttechnologischer Methoden liegt im Textmining[8]. Für das hier beschriebene Vorhaben sind Verfahren des Textmining auf mehreren Ebenen von Relevanz: für die Terminologieextraktion im Ontologieaufbau (Buitelaar et al. 2005; Maedche, Volz 2001), für das Auffinden von Belegstellen für bereits in der Datenbank enthaltene Signalpfade, sowie für die Extraktion von Fakten aus Texten (Dickerson et al. 2003).

2.7 Netzgestützte Wissenschaftskommunikation

Die derzeit hauptsächlich genutzten Formen der Wissenschaftskommunikation sind in mancherlei Hinsicht in eine Krise geraten. Wissenschaftliche Zeitschriften können – besonders in aktiven Forschungsgebieten - mit ihrem Erscheinungsrhythmus immer weniger der schnellen Entwicklung folgen. Ihr ständig steigender Preis stellt wiederum die wissenschaftlichen Bibliotheken vor große Probleme. Als Reaktion wurden Verfahren vorgeschlagen, wie Wissen mit Hilfe der Kommunikationsmöglichkeiten des Internet schneller und flexibler zur Verfügung gestellt werden kann, als dies mit den traditionellen Publikationsorganen möglich sein kann[9]. Diese neuen Ansätze des Knowledge Managements (Kuhlen 2003) beruhen auf dem Aufbau von Wissensportalen durch egalitäre Beteiligung einer interessierten Gruppe[10]. Gegenstand der Arbeit können Enzyklopädien (Stallman 2005) sein oder spezialisiertere Forschungsgebiete. Entscheidend ist, dass einerseits Anreizsysteme

---

[7] Hier sind über die Projektlaufzeit die Ergebnisse der Arbeitsgruppen des W3C zu Beschreibung und Interaktion von Web-Services zu berücksichtigen (http://www.w3c.org/2002/ws/, zitiert am 28.1.05). Besondere Relevanz haben die Ergebnisse der im Kompetenznetzwerk „Neue Dienste, Standardisierung, Metadaten“ vom BMBF geförderten Projekte, insbesondere des Teilprojekts „Generische und komponentenbasierte wissenschaftliche Portale“.

[8] Einen Überblick über die Methoden bietet (Hotho et al. 2005).

[9] Hier sind auch Internetportale wie etwa vascoda (www.vascoda.de/, zitiert am 28.1.05. vgl. Neuroth; Pianos 02) zu berücksichtigen.

[10] Diese Forschung wird unter dem Titel *K3 – Wissensmanagement über kooperative verteilte Formen* über den Projektträger im DRL NMB+F vom BMBF gefördert.

zur Teilnahme an derartigen Projekten auffordern sollen, gleichzeitig aber Qualitätsindikatoren den Stellenwert von Beiträgen verdeutlichen müssen (Semar 2004). Als Folge verschwimmen – hinsichtlich der Zugänglichkeit, nicht der Qualitätseinschätzung – die Grenzen zwischen einer begutachteten Fachpublikation und grauer Literatur.

Die Integration der Fachinformation in einem Informationsportal erleichtert zudem, einen Überblick über den aktuellen Forschungsstand zu gewinnen.

3. Vorgehensweise

Eingangs wurden grundsätzlich zwei Problemzonen in der Erschließung und Kommunikation wissenschaftlicher Ergebnisse über Signalpfade identifiziert:

1. Die Erschließungstiefe bisheriger Signalpfaddatenbanken ist nicht ausreichend. Sie orientieren sich primär an intrazellulären Prozessen, geben wenig oder keine Auskunft über funktionale Zusammenhänge und berücksichtigen nicht die Kompartimente, in denen diese Prozesse ablaufen. Des weitern sind keine Instrumente für die Integration klinischer Forschungsergebnisse vorhanden.
2. Die Wissenschaftskommunikation in der Erforschung von Signalpfaden verläuft in vielen Bereichen ineffizient. Als primäres Kommunikationsmittel stehen wissenschaftliche Zeitschriften mit ihrem hohen Qualitätsniveau aber auch langen Vorlaufzeiten zur Verfügung. Ein offene Kommunikationsplattform, die es erlaubt, wissenschaftliche Hypothesen zur Debatte zu stellen könnte hier Abhilfe schaffen.

Aus diesen zwei recht allgemein beschriebenen Problemfeldern lassen sich folgende konkretere Arbeitsbereiche mit jeweils eigenen Methoden ableiten:

1. Formale Grundlagen: Teil des Projekts ist der Entwurf neuer Ontologien und Annotationsschemata. Zu klären ist, wie begriffliche und mereologische Relationen ausgedrückt werden, welche Annotationsschemata für biochemische, physiologische bzw. klinische Phänomene benötigt werden und wie diese untereinander in Beziehung zu setzen sind. Dabei werden auch Inferenzschemata zu definieren sein.
2. Informationsbedarfsanalyse: Vor konkreten Systementwürfen muss eine genaue Analyse des durch diese System zu befriedigenden Informationsbedarfs stehen (Kluck 1997). Durch Umfragen oder Nutzungsstatistiken kann so auf die Konzeptualisierung des Informationsbedarfs durch die Nutzer geschlossen werden, so dass die Strategien der Inhalterschließung und die angebotenen Recherchemöglichkeiten diesem Bedarf angepasst werden können.
3. Akzeptanzanalyse: In Disziplinen, deren Publikationsstandards sehr stark von Impact-Faktoren geprägt sind, werden alternative, informellere Publikationsformen nur schwer Akzeptanz finden. Um Fehlentwicklungen zu vermeiden sind frühzeitig Akzeptanzuntersuchungen durchzuführen. Durch das Angebot vergleichbarer Qualitätsstandards – etwa durch Einsatz von angepassten bibliometrischen Verfahren für das Web (Ball, Tunger 2005) – könnten Widerstände überwunden werden.
4. Korpusanalyse: Der Aufbau der Ontologie erfolgt durch Extraktion von Terminologie und terminologischen Relationen aus einem angemessen ausgewählten Korpus, das zunächst zusammenzustellen ist. Dabei soll soweit als möglich, sowohl was die Analysetools angeht, als auch hinsichtlich der Korpora, auf schon bestehende Ressourcen zurückgegriffen werden (Morgan 2004).
5. Visualisierungsverfahren: Die bestehenden Ansätze zur Visualisierung von Signalpfaden sind zu sichten und so erweitern, dass rein funktionale Zusammenhänge

durch Lokalisierungsinformation (Kompartiment) ergänzt wird. Dabei wird, soweit möglich, nicht nur auf die Bildsprache sondern auch auf Abbildungskorpora aus anatomischen Lehrmaterialen zurückgegriffen.

6. Konstruktion eines Informationssystems: Gegenstand des Projekts sind in erster Linie die Entwicklung neuer Erschließungs- und Recherchemöglichkeiten für Signalpfade sowie die Eröffnung neuer Kommunikationswege. Für die Software wird daher eine offene Architektur gewählt, welche die Integration bestehender Komponenten erlaubt. Ziel ist es, die vorgeschlagenen Ansätze an einem experimentellen aber operationalen Prototypen evaluieren zu können.

Die Aufteilung in eher empirisch und eher informationsmethodisch dominierte Fragestellungen entspricht weitgehend der Arbeitsteilung zwischen den Kooperationspartnern aus Medizin und Informationswissenschaft.

Obschon das Projekt hinsichtlich zentraler Fragen Grundlagenforschung leistet – etwa in der Ausgestaltung von Ontologie und Annotationsschemata – ist es dennoch strukturell so angelegt, dass auch der organisatorische Rahmen für eine nachhaltige Einführung eines derartigen Informationssystems von vornherein geschaffen wird. Ein umfänglicher, international besetzter wissenschaftlicher Beirat schafft die Voraussetzung für die Etablierung von Qualitätsstandards. Frühzeitige Einbindung von Verlagen erhöht die Wahrscheinlichkeit, dass das zu konzipierende Informationssystem mit seinen Formen der Wissenschaftskommunikation harmonisch in das bestehende System integriert wird. Weiterhin ist der langfristige Betrieb zu gewährleisten, für den eine akademische Einrichtung nicht eintreten kann.

4. Projektziele und Stand der Arbeiten

Hauptziel des Projektes ist es, ein neuartiges Informationssystem über Signalpfade nachhaltig zu etablieren. Unabhängig von diesem Globalziel können Teilziele definiert werden, die Ergebnisse von eigenem Wert mit jeweils individuellen Nutzungsmöglichkeiten erzielen werden.

1. Vorschläge für erweiterte Ontologien und Annotationsschemata sind als Standardisierungsvorschläge den zuständigen Gremien vorzulegen und somit dauerhaft zu etablieren.
2. Erschlossene Korpora werden der Fachöffentlichkeit zur Verfügung gestellt.
3. Empirische Untersuchungen zum Informationsbedarf und zur Akzeptanz wissenschaftlicher Kommunikationsformen stellen einen eigenständigen wissenschaftlichen Wert dar.
4. Softwarekomponenten können unter einer offenen Lizenz zur Verfügung gestellt werden.

Zur Zeit befindet sich das Projekt in einer frühen Phase. Vorstudien zu Informationsbedarfs- und Akzeptanzanalyse sind mit einer kleinen Expertengruppe durchgeführt worden. Ihr Ergebnis war ermutigend, bedarf aber noch der Absicherung durch umfangreichere Befragungen. Als nächste Schritte sind die Auswahl von Korpora und die formale Grundlegung der zu erstellenden Ontologie vorgesehen.

Literaturangaben

Ball, Rafael; Tunger, Dirk (2005) Bibliometrische Analysen – Daten, Fakten und Methoden. Schriften des Forschungszentrums Jülich, Reihe Bibliothek, Bd. 12.

Bard, Jonathan B.L.; Rhee, Seung Y. (2004) Ontologies in Biology. Design, Applications and Future Challanges. Nature Reviews, Genetics, Bd. 5, March 2004, S. 213-222.

Battistella, E.; de Souza, J.C.G.; Ferreira, R.A.; Vieira, R.; Mombach, J.C.M.; Lemke, N. (2004) Bioinformatics: A Growing Field for Ontologies. Workshop on Ontologies and their Applications. 28.9.2004, Sao Luis, Brasilien. http://www.ws.onto.ufal.br/Papers/Battistella.pdf (20.1.2005).

Buitelaar, P.; Cimiano, P.; Magnini, B. (2005) Ontology Learning from Text: Methods, Evaluation and Applications. IOS Press.

Becker, M.Y.; Rojas, I. (2001) A graph layout algorithm for drawing metabolic pathways. In. Bioinformatics, Bd. 17, Nr. 5, S. 461-467.

Deville, Y.; Gilbert, D.; Helden, J.; Wodak, S. (2003) An Overview of Data Models for the Analysis of Biochemical Pathways. In: Briefings in Bioinformatics, Bd. 4, Nr. 3, S. 246-259.

Dickerson, J.A.; Berleant, D.; Cox, Z; Qi, W.; Ashlock, D.; Wurtele, E. (2003) Creating Metabolic Network Models using Text Mining and Expert Knowledge. http://www.public.iastate.edu/~mash/publications/dickerson03b.pdf, zitiert am 17.7. 2005

Ding, Y.; Foo, S. (2002a) Ontology research and development. Part 1: A Review of Ontology Generation. Journal of Information Science, Bd. 28, Nr. 2, S. 123-136.

Ding, Y ; Foo, S. (2002b) Ontology research and development. Part 2: A Review of Ontology Mapping and Evolving. Journal of Information Science, Bd. 28, Nr. 5, S. 375-388.

Fan, J.; Gao, Y.; Luo, H. (2004) Multi-level annotation of natural scenes using dominant image components and semantic concepts. In Proc. of the 12th annual ACM international conference on Multimedia. S. 540-547.

Goesman, Alexander; Haubrock, Martin; Meyer, Folker; Kalinowski, Jörn; Giegerich, ,Robert (2002) Pathfinder: Reconstruction and Dynamic Visualization of Metabolic Pathways. In Bioinformatics, Bd. 18, Nr. 1, S. 124-129.

Ghandeharizadeh, S.; Sommers, F.; Joisher, K.; Alwagait, E. (2002) A document as a web service: Two complementary frameworks. In: Chaudhri, A.B. (Hrsg.) et al. XML-based data management and multimedia engineering. Springer, Berlin, S. 450-461.

Giogno, Rosalba; Shasha, Dennis (2002) GraphGrep: A Fast and Universal Method for Querying Graphs. In Proceeding of the International Conference in Pattern recognition (ICPR), Quebec, Canada, August 2002. http://www.cs.nyu.edu/shasha/papers/graphgrep/icpr2002.pdf, zitiert am 18.7.2005

Hotho, Andreas; Nürnberger, Andreas; Paaß, Andreas (2005) A Brief Survey on Text Mining. In LDV Forum, Bd. 20, Nr. 1, S. 19-62.

Joshi-Tope, G.; Gillespie, M.; Vastrik I, D'Eustachio, P.; Schmidt, E.; de Bono, B., Jassal, B.; Gopinath, G.R.; Wu, G.R.; Matthews, L.; Lewis, S.; Birney, E. Stein, L. (2005) Reactome: a knowledgebase of biological pathways. In Nucleic Acids Res. Bd. 33, Nr. 1, Database Issue. S. 428-432.

Kanne, C.-C.; Schreiber, F.; Trümbach, D. (1999) Electronic Biochemical Pathways. In Kratochvíl, J. (Hrsg.), Graph Drawing. Proc. 7th International Symposium. GD'99, Stirín Castle, Czech Republic, September 1999, S. 418.

Karopka, T.; Scheel , T.; Bansemer, S.; Glass, A. (2004) Automatic construction of gene relation networks using text mining and gene expression data. In: Med Inform Internet Med., Bd. 29, Nr 2, S. 169-183.

Karp, P.D.; Paley, S.M.; Krieger, C.J.; Zhang, P. (2004) An Evidence Ontology for Use in Pathway/Genome Databases. Pacific Symposium on Biocomputing. S. 190-201, (http://helix-web.stanford.edu/psb04/karp.pdf, zitiert am 28.1.05).

Michael Kluck (1997): Methoden der Informationsanalyse. In: Buder, M.; Rehfeld, W.; Seeger, TH.; Strauch, D. (Hrsg.): Grundlagen der praktischen Information und Dokumentation. München et al.: K.G. Saur, S. 795-821

Koike, A.; Niwa, Y.; Takagi, T (2004) Automatic extraction of gene/protein biological functions from biomedical text. In: Bioinformatics, epub.

Krieger, Cynthia J.; Zhang, Pelfen; Müller, Lukas A.; Wang, Alfred; Paley, Suzanne; Arnaud, Martha; Pick, John; Rhee, Seung Y; Karp, Peter D. (2004) Nucleic Acids Research, Bd. 32, Nr. 10.

Kuhlen, R. (2003) Change of Paradigm in Knowledge Management – Framework for the Collaborative Production and Exchange of Knowledge. In Hobohm, H.-C. (Hrsg.) Knowledge Management – and Asset for Libraries and Librarians. Collected Papers from LIS Professionals.

Lewis, S.E. (2005) Gene Ontology: Looking Backwards and Forwards. In: Genome Biol. Bd. 5, Nr. 1.

Lomax, J.; McCray, A. (2004) Mapping the Gene Ontology into the Unified Medical Language System. Comparative and Functional Genomics, Bd. 5, Nr. 4, S. 354-361.

Maedche, Alexander; Voltz, Raphael (2001) The Ontology Extraction & Maintainance Framework Text-To-Onto. In ProcWorkshop on Integrating Data Mining and Knowledge Management http://cui.unige.ch/~hilario/icdm-01/DM-KM-Final/Volz.pdf, zitiert am 18.7.2005.

Morgan, Alex (2004) BioNLP-Resources. http://www.tufts.edu/~amorga02/bcresources.html, zitiert am 17.07..2005.

Müller, C.; Strube, M. (2003) Multi-Level Annotation in MMAX. In Proc. of the 4th SigDial Workshop on Discourse and Dialogue, Sapporo, 5-6 July 2003, S. 198-207.

Neuroth, H.; Pianos, T. (2003) VASCODA: A German Scientific Portal for Cross-Searching Distributed Digital Resource Collections. Proc. 7th Int. Conf. on Research and Advanced Technology for Digital Libraries, S. 257-262.

Paek, E.; Park, J.; Lee, K.J. (2004) Multi-layered representation for cell signaling pathways. In Mol Cell Proteomics. Bd. 3, Nr. 10, S. 1009-22.

Pinter, Ron Y.; Rokhlenko, Oleg; Yeger-Loten, Esti; Ziv-Ukelson, Michal (2005) Alignment of Metabolic Pathways. In Bioinformatics, zur Publikation angenommen.

Rahman S.A.; Advani, P.; Schunk, R.; Schrader, R.; Schomburg, D. (2004) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). In: Bioinformatics. Nov 30, epub.

Rojdestvenski, I. (2003) Metabolic pathways in three dimensions. Bioinformatics. Bd. 19, Nr. 18, S. 2436-2441.

Sasaki, F. (2004) Sekundary Information Structuring – A Methodology for the Vertical Interrelation of Information Resources. In Proc. of Extreme Markup Languages, Montréal, Kanada.

Schacherer, Frank (2001) An object-oriented database for the compilation of signal transduction pathways. Dissertation, Technische Universität Carolo-Wilhelmina Braunschweig.

Schreiber, F. (2002) High Quality Visualization of Biochemical Pathways in BioPath. In Silico Biology, Bd. 2, Nr. 6.

Semar, Wolfgang (2004) Incentive Systems in Knowledge Management to Support Cooperative Distributed Forms of Creating and Acquiring Knowledge. In: Arabnia, Hamid; et al. (Hg.): Proceedings of the International Conference on Information and Knowledge Engineering - IKE'04. Las Vegas: CSREA Press, S. 406 - 411

Sivakumaran Sivakumaran, S.; Hariharaputaran, S.; Mishra, J.; Bhalla, U.S. (2003) The Database of Quantitative Cellular Signalling: Management and Analysis of Chemical Kinetic Models of Signalling Networks. Bioinformatics, Bd. 19. Nr. 3, S. 408-415.

Smith, Barry; Ceusters, Werner; Klagges, Bert; Köhler, Jacob; Kumarm, Anand; Lomax, Jane; Mungall, Chris; Neuhaus, Fabian; Rector, Alan L.; Rosse, Cornelius (2005a) Relations in Biomedical Ontologies. In Genome Biology, Vol. 6, Nr. 5, http://genomebiology.com/2005/6/5/R46, zitiert am 17.7.2005.

Smith, Barry; Mehino, Jose L.V.; Schulz, Stefan; Kumar, Anand; Rosse, Cornelius (2005b) Anatomical Information Science. http://ontology.buffalo.edu/anatomy_GIS/FMA-AIS.pdf, zitiert am 17.7.2005.

Sohler, Florian; Zimmer, Ralf (2004) Identifying Active Transcription Factors and Kinases from Expression Data using Pathway Queries. In Bioinformatics, Bd. 20, 1517 – 1521. http://www.bio.ifi.lmu.de/mitarbeiter/sohler/eccb05.pdf, zitiert am 18.7.2005.

Stallman, R. (2005) The Free Universal Encyclopedia and Learning Resource. [http://www.gnu.org/encyclopedia/free-encyclopedia.html], zitiert am 21.1.2005

Takai-Igarashi, Takako; Mizoguchi, Riichiro (2003) Cell Signaling Network Ontology. http://www.bioinfo.de/isb/2003/04/0008/main.html, zitiert am 17.7.2005.

Teich E.& S.Hansen (2001) Towards an integrated representation of multiple layers of linguistic annotation in multilingual corpora. In Online Proceedings of Computing Arts 2001: Digital Resources for Research in the Humanities. Sydney, http://www.coli.uni-sb.de/~hansen/teichhansen-final.pdf, zitiert am 28.1.2005

Vargas-Vera, M.; Motta, E.; Domingue, J.; Lanzoni, M.; Stutt, A.; Ciravegna, F. (2002) MnM: Ontology driven semi-automatic and automatic support for semantic markup. In: Gómez-Pérez, A. (Hrsg) et al., Knowledge engineering and knowledge management. Ontologies and the semantic web. 13th international conference, EKAW 2002, Sigüenza, Spain, October 1-4, 2002. Proceedings. Berlin: Springer. Lect. Notes Comput. Sci. 2473, S. 379-391.