



Hochschule Rosenheim
University of Applied Sciences



HOLZBAU DER ZUKUNFT

**TP 21 und 22 Integriertes Relationales
Informationssystem
für den Holzbau
- IRIS -**



Hochschule Rosenheim
University of Applied Sciences



Univ. Prof. DI Hermann Kaufmann
Technische Universität München,
Lehrstuhl Baukonstruktion, Fach-
gebiet Holzbau

Prof. Dr. Hartmut Ernst
Hochschule Rosenheim
Fachbereich Informatik

TEILPROJEKTE 21 UND 22

**Integriertes Relationales Informationssystem
für den Holzbau
- IRIS -**



Univ. Prof. DI Hermann Kaufmann
Technische Universität München,
Lehrstuhl Baukonstruktion, Fach-
gebiet Holzbau

Hochschule **Rosenheim**
University of Applied Sciences



Prof. Dr. Hartmut Ernst
Hochschule Rosenheim
Fachbereich Informatik

1 Zusammenfassung

Innovation und Kostensenkung sind zwei wesentliche Faktoren für die Zukunftsfähigkeit im Bereich des Holz- und Holzhausbaus, sowohl der industriell strukturierten Unternehmen, als auch der kleinen und mittleren Unternehmen (KMU). Ebenso bestimmende Faktoren sind die Markt-, Produkt- und Kundenorientierung. Der Aktionsraum der Unternehmen verlagert sich auch im Holz- und Holzhausbau weg von den regionalen bzw. nationalen Märkten hin zu den internationalen Märkten.

Zur Bewältigung der hieraus resultierenden vielfältigen Aufgaben und Anforderung sind zeitnahe und verlässliche Daten und Informationen erforderlich. Das stellt die zielgerichtete Beschaffung, Bereitstellung und Verarbeitung von Informationen und Daten als einen wettbewerbsentscheidenden Faktor in den Mittelpunkt unternehmerischen Handelns.

Für die Informationsbeschaffung zu forschungsrelevanten Themen oder für die Entwicklung neuer Produkte und Technologien stehen dem Anwender im Bauwesen allgemein, so auch im Holzbau, gegenwärtig eine Vielzahl von „unkoordinierten Insellösungen“ zur Verfügung. Ein übergeordnetes Dienstleistungssystem, welches für Forschungs- und Entwicklungsprozesse im Holzbau den Abruf strukturierter Daten und Informationen in digitaler Form ermöglicht, fehlt bislang. Das Projektziel wurde daher wie folgt formuliert:

„Es ist ein übergeordnetes Dienstleistungssystem zu entwickeln, das für Forschungs- und Entwicklungsprozesse im Holzbau den Abruf strukturierter Daten und Informationen ermöglicht. Des Weiteren dient das Dienstleistungssystem als Schnittstelle zu vorhandenen Daten- und Informationsbeständen.“

Der Einstieg in die Konzeption erfolgte mit mehreren Brainstormings innerhalb des Projektteams und in Gesprächen mit Projektpartnern, Unternehmen und Verbänden, um Fragen zu klären und Bedürfnisse an ein Informationssystem zu ermitteln. Wichtigstes Ergebnis aller Gespräche war, dass die Realisierung einer Plattform für den Informationsaustausch in der Tat notwendig ist.

Da eine Einrichtung nicht alle Themengebiete im Holzbau fachlich und personell abdecken kann und um die Verantwortung für die Qualität und den wissenschaftlichen Anspruch der Daten und Informationen sowie den finanziellen und zeitlichen Aufwand auf mehrere Schultern zu verteilen, erschien ein Kooperationspartner-Konzept sinnvoll. Dieses Konzept sieht die Einbindung von Experten, Hochschulen und Instituten als gleichberechtigte Partner vor. Jeder Kooperationspartner ist Pate für ein Themengebiet und steht für dessen inhaltliche Ausgestaltung und die redaktionelle Unterstützung zur Verfügung. Dies bedeutet, dass die Kooperationspartner inhaltliche Beiträge liefern, die dann im Informationssystem entsprechend aufbereitet und dem Benutzer zugänglich gemacht werden.

Mit IRIS – für „Integriertes Relationales Informationssystem“ – soll ein Beitrag zur Informationsvermittlung und Wissensmanagement für alle Beteiligten im Holz- und Holzhausbau, nachfolgend Holzbau genannt, geleistet werden. Das bedeutet, dass hauptsächlich Themenbereiche, die in (unmittelbarem) Zusammenhang mit dem Holzbau stehen, betrachtet werden.

Das Dienstleistungssystem IRIS soll keine vorhandene Informationsplattform oder Datenbank ersetzen, sondern vor allem eine Schnittstelle zwischen den

einzelnen „Insellösungen“ bieten. Sowohl die Information gebenden Kooperationspartner als auch die Systembetreiber profitieren von IRIS.

Konzipiert als ein sich selbst entwickelndes, „lernendes“ internetbasiertes Informationssystem, wurde mit IRIS innerhalb der Projektlaufzeit von 32 Monaten in drei Phasen eine Webplattform zur branchenübergreifenden Integration von fachlich relevanten Informationen und Dokumenten geschaffen. Seit August 2007 läuft die Anwendung als Testversion auf einem projekteigenen HP Proliant Server und steht unter der Adresse <http://iris.fh-rosenheim.de> im Internet zur Verfügung.

Für die Abdeckung des umfassenden Informationsbedürfnisses der Anwender gibt es eine einfach zu benutzende Volltextsuche ähnlich einer üblichen Suchmaschine wie Google, ein semantisches Netz mit verwandten Begriffen zur visuellen Unterstützung bei der Wahl geeigneter Suchbegriffe und eine Enzyklopädie mit Artikeln rund um das Thema Holzbau.

Die Auswahl von Inhalten für die Suchfunktion beschränkt sich auf ausgewählte Quellen im Internet sowie von Kooperationspartnern bereitgestellte Dokumente. Diese Beschränkung auf hochwertige Inhalte bietet dem Suchenden einen großen Mehrwert gegenüber herkömmlichen Internet-Suchmaschinen, bei denen die Qualität der Treffer unklar ist und stark variiert. Mit folgenden möglichen Partnern wurden Kooperationsgespräche zur Bereitstellung von Inhalten für IRIS durchgeführt:

- DIN/Beuth-Verlag, Berlin
- Deutsches Patent- und Markenamt (DPMA), und europäisches Patentamt (EPO), beide München
- Informationszentrum Raum und Bau (IRB), Stuttgart
- Technische Informationsbibliothek (TIB), Hannover
- Holzabsatzfonds / Informationsdienst-Holz / infoholz.de (IDH) Bonn
- Deutsche Gesellschaft für Holzforschung (DGfH), München
- Institut für Fenstertechnik (ift), Rosenheim.

Die DGfH ist der wichtigste Kooperationspartner. Neben der inhaltlichen Unterstützung mit Forschungsberichten und Dokumenten soll sie die Informationsplattform nach Projektende weiterführen. Die DGfH unterstützte das Projekt aktiv bei der Anbahnung von Kontakten mit potentiellen Kooperationspartnern.

Das semantische Netz unterstützt den Benutzer insbesondere beim Suchvorgang. Es macht Zusammenhänge sichtbar und erweitert die herkömmlichen Suchmethoden (Volltext, Schlagwort usw.). Daraus ergeben sich unter anderem folgende Vorteile:

- Das Navigieren durch das semantische Netz hilft dem Benutzer, das richtige Vokabular zur Fragebildung zu finden und somit das passende Ergebnis effektiv und effizient zu erlangen
- Der Benutzer kann sich ein Themengebiet sowohl begrifflich als auch hinsichtlich der Zusammenhänge schnell erarbeiten. Möchte er einen Punkt vertiefen, so kann er unmittelbar in passenden Quellen nachlesen
- Eine komfortable visuelle Navigation per Mausclick schafft zudem einen guten Überblick und macht die Suche schnell

Im Rahmen des Projektes war eine holzbauspezifische Systematik zu erarbeiten. Ein von den Projektbearbeitern entwickeltes Grundkonzept wurde in Zusammenarbeit mit dem Institut für angewandte Technologien im Holzbau e. V. (IATH), Übersee, verfeinert. Die holzbauspezifische Systematik wurde

vollständig im semantischen Netz der Anwendung abgebildet. Die begriffliche Vernetzung ist jedoch nicht auf das Thema Holzbau beschränkt, sondern kann ebenso für andere Wissensgebiete genutzt werden.

Das semantische Netz kann sich selbst weiterentwickeln und neue Begriffe und Verknüpfungen „dazulernen“ oder auch wieder „vergessen“. Dazu werden Nutzeranfragen über einen gewissen Zeitraum ausgewertet. Der eigentliche Lernprozess ist dann manuell von einem berechtigten Redakteur anzustoßen. Das Ergebnis des Lernvorgangs verbessert sich mit dem Umfang an auswertbarer Nutzerinteraktion.

IRIS selbst ist als Webapplikation auf Basis von Java EE (Java Plattform, Enterprise Edition) gemäß der MVC2-Architektur (Model-View-Controller-Architektur für Webanwendungen) realisiert. Das Durchlaufen des Suchraumes („Crawling“) sowie die Suchfunktion der Oberfläche erfolgt mit einer modifizierten und erweiterten Version der Open-Source Software Regain. Für die Indexierung von Inhalten kommt die Open-Source Bibliothek Lucene zum Einsatz. Die Enzyklopädie wurde als integriertes Wiki-System mit einer angepassten Version von JSPWiki realisiert. Der gesamte Quellcode ist in Java geschrieben. Die Präsentation wird von Java Server Pages (JSPs) erledigt. Für die inhaltliche Redaktion und Administration des Systems wurden eine geeignete Weboberfläche und ein mehrstufiges System von Benutzerrechten in IRIS implementiert. Die Applikation läuft derzeit auf dem Jakarta Tomcat Servlet Container unter SUSE Linux Enterprise Edition.

Die Software-Entwicklung wurde im ersten Quartal 2008 im Wesentlichen abgeschlossen. Seither werden daran vor allem Verbesserungen, Fehlerbehebungen und kleinere Funktionserweiterungen zur Abrundung des Systems vorgenommen. Aus Sicht des Projektteams werden die wesentlichen Forderungen, die im Rahmen dieses Forschungsprojektes an das System gestellt werden, bewältigt. Gleichwohl ist es weiter ausbaufähig.

Seit Mitte August 2007 ist IRIS als Testversion im Internet unter der Adresse <http://iris.fh-rosenheim.de> verfügbar. Seither werden Praxiserfahrungen im Betrieb gesammelt, z. B. bzgl. des Einspielens von Patches und Updates. Aufgrund der bedauerlicherweise zuletzt noch immer verhaltenen Resonanz seitens der Zielgruppen konnten jedoch noch kaum Erkenntnisse hinsichtlich der Effektivität des „Lernenden Systems“ – welches auf ein gewisses Mindestmaß an Nutzerinteraktion angewiesen ist – oder der Bewältigung von Spitzenlasten gewonnen werden.

Um IRIS nach dem Ende der Projektlaufzeit in die Praxis zu überführen, wurde im März 2008 beim Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) im Rahmen des Programms „FIT für den Wissenswettbewerb – Transferaktivitäten“ unter dem Titel „Ontologisch-semantisches integriertes Relationales Informationssystem – OSIRIS“ eine Projektskizze eingereicht. Der Schwerpunkt dieses anvisierten Nachfolgeprojektes ist der Transfer der bisher geleisteten Arbeiten und Ergebnisse in die Praxis. Zum Zeitpunkt der Erstellung dieses Berichtes gab es von Seiten des DLR noch keine Stellungnahme zu einer möglichen Förderung von OSIRIS.

2 Summary

Innovation and cost reduction are two essential factors for the sustainability in the field of timber construction and the construction of wooden houses as well for industrially structured companies as for small and medium-sized enterprises (SME). Likewise determining factors are the orientation on markets, on products and on customers. The focus of the enterprises of the timber construction and the construction of wooden houses, also, relocates from the regional and national markets to the international markets.

Contemporary and reliable data and information are necessary for the coverage of the resulting various tasks and requirements. This shifts the goal-oriented retrieval, supply and handling of information and data to the center of entrepreneurial action.

Currently, a multitude of "uncoordinated stand-alone solutions" for the retrieval of information about topics pertaining to research or for the development of new products and technologies is available to users of the construction engineering sector in general as well as for the timber construction sector. There is no superordinated service system that provides the retrieval of structured data and information in an electronic format for research and development processes in the timber construction sector so far. The project goal was therefore stated as follows:

"A superordinated service system has to be developed which enables the retrieval of structured data and information in an electronic format for research and development processes in the timber construction sector. Furthermore, the service system serves as interface to existing databases and stocks of information."

For starting the design several brainstormings were held within the project team and in interviews with project partners, companies and associations to clarify questions and to keep hold of the requirements of an information system. The most important result of all interviews was that the realization of a platform for the exchange of information is indeed necessary.

As a single institution cannot cover all topics of the timber construction technically and personally and to pool the responsibility for the quality and scientific standard of the data and information as well as the expenses and time requirements a co-operation partner concept seemed to make sense. According to the concept experts, universities and institutes shall be involved as equals. Every co-operation partner sponsors one topic and provides its contents design and editorial support. That means the co-operation partners will provide content that will then be edited in the information system and made accessible to the users.

IRIS – the Integrated Relational Information System – shall be a contribution to information brokering and knowledge sharing for all parties of the timber construction sector and the construction of wooden houses sector, hereinafter referred to as "timber construction". That means that mainly topics (directly) linked to timber construction will be considered.

The information system IRIS shall not replace existing information platforms or databases but provide an interface between the single stand-alone solutions. Both the information providing co-operations partners and the system operators benefit from IRIS.

As designed as a self-developing, "learning" internet based information system, IRIS was created within the project duration of 32 months in three phases being a web platform for the intersectoral integration of technically relevant information and documents. Since August 2007 the application runs as a test version on an HP Proliant server owned by the project and is accessible through the internet via the address <http://iris.fh-rosenheim.de>.

For the coverage of the comprehensive information demands of the users there are a simple to use full-text search function similar to common search engines like Google, a semantic net with related terms to visually assist in finding appropriate search terms and an encyclopedia with articles related to the topic timber construction.

The assortment of content for the search function is limited to selected sources of the internet and documents provided by the co-operation partners. That limitation to high-quality content is a big surplus to the user compared to usual internet search engines which quality of the hits is unknown and diverging widely. Cooperation talks about the provision of content for IRIS were held with the following possible partners:

- DIN/Beuth-Verlag, Berlin
- German Patent and Trade Mark Office, and European Patent Office (EPO), both Munich
- Information Centre for Planning and Building (IRB), Stuttgart
- German National Library of Science and Technology (TIB), Hanover
- Holzabsatzfonds / Informationsdienst-holz / infoholz.de (IDH) Bonn
- German Society for Wood Research (DGfH), Munich
- Institut für Fenstertechnik (ift), Rosenheim.

The DGfH is the most important co-operation partner. Besides the support regarding content with research reports and documents it shall continue operating the information platform after the project has ended. The DGfH actively supported the project in the initiation of contacts to potential co-operation partners.

The semantic net assists the user especially during the search operation. It visualizes coherences and extends the common search techniques (full-text, keyword etc.). This results in the following advantages:

- Navigating through the semantic net helps the user find the proper vocabulary to create search requests and thus get the matching result in an effective and efficient way
- The user can quickly acquire a topic conceptually as well as regarding the coherences. If he is willing to deepen an item he can directly look up in matching sources
- A comfortable visual navigation via mouse click creates a good general overview and makes the search fast

Within the project a timber construction specific taxonomy had to worked out. A basic concept developed by the project employees was improved in co-operation with the Institut für angewandte Technologien im Holzbau e. V. (IATH), Übersee. The timber construction specific taxonomy became fully mapped in the semantic net of the application. The cross-linking of terms is not limited to the topic timber construction though, but can be applied to other fields of knowledge as well.

The semantic net is able to improve itself by "learning" new terms and connections or as well "forgetting" them. For that user requests from a certain

time span are evaluated. The actual learning task has to be started by an authorized editor. The result of the learning operation improves according to the amount of evaluable user interaction.

IRIS is realized as a web application based on Java EE (Java platform, Enterprise Edition) according to the MVC2-architecture (Model-View-Controller-architecture for web applications). Running through the search space ("Crawling") as well as the search functionality on the user interface is done by a modified and extended version the open-source software Regain. For indexing of content the open-source library Lucene is used. The encyclopedia is realized as an integrated Wiki system using a customized version of JSPWiki. The whole source code is written in Java. The presentation is done by Java Server Pages (JSPs). For the content-related editorial work and administration of the system a suitable web interface and a multi-level system of user rights were implemented in IRIS. The application is currently running on a Jakarta Tomcat servlet container under SUSE Linux Enterprise Edition.

The software development was largely completed in the first quarter in 2008. Since then mainly improvements, error fixes and minor functionality extensions to round off the system are performed. According to the project team, the essential requirements of the system defined within this research project are fulfilled. Nonetheless, it could still be expanded.

Since august 2007 IRIS is available as a test version on the internet under the address <http://iris.fh-rosenheim.de>. Since then, practical experience is collected, e. g. about applying patches and updates. Because of the unfortunately still poor feedback by the target groups hardly any results could be gained yet about the effectiveness of the "learning system" which depends on some minimum user interaction or dealing with peak workloads.

To practically apply IRIS after the end of the project duration a project proposal titled "Ontologisch-semantisches integriertes Relationales Informationssystem – OSIRIS" (Ontological-semantic integrated relational information system) was submitted to the German Aerospace Center (DLR) under the program "Fit for Competition in the Knowledge Society" in march 2008. The focus of that proposed succeeding project is the transfer of the results gathered so far and the work performed till now into practice. By the time when this report is written there is not yet a response from the DLR concerning a possible funding of OSIRIS.

3	Inhaltsverzeichnis	5
1	Zusammenfassung.....	5
2	Summary.....	8
3	Inhaltsverzeichnis	11
4	Tabellenverzeichnis.....	14
5	Abbildungsverzeichnis.....	14
6	Projektbeschreibung.....	16
6.1	Ausgangslage	16
6.2	Projektziel.....	18
6.3	Projektorganisation und Projektterminplan	19
6.4	Ableich mit dem Projektantrag	20
7	Bestandsaufnahme	21
7.1	Beschreibung des Branchenumfelds – Strukturdaten.....	21
7.2	IuK – Zugangsmöglichkeiten	22
7.3	Bestandsaufnahme und Analyse bestehender Datenbank- und Informationssystemlösungen	25
8	Inhaltliche Konzeption und Realisierung	26
8.1	Motivation	26
8.2	Ideensammlung zu IRIS	27
8.3	Projektgestaltung.....	29
8.4	Alleinstellungsmerkmale von IRIS.....	30
8.5	Vorstellung des Kooperationspartner-Konzepts	30
8.6	Inhaltliche Themengebiete	31
8.7	Zielgruppen und Anwender	32
8.8	Lastenheft und Pflichtenheft	33
8.9	Aufbau der Informationsplattform	33
8.10	Daten, Dokumente und Artikel.....	34
8.11	Zum Projektabschluss realisierte Kooperationen	35
8.12	Qualitätssicherung	37
8.13	Öffentlichkeitsarbeit.....	37
9	Fachsystematik – Holzbauspezifische Systematik.....	38
9.1	Begriffsdefinitionen	38
9.1.1	Möglichkeiten der Strukturierung	39
9.1.2	Vorhandene Systematiken	40
9.2	Fachsystematik Holzbau	41
9.3	Semantisches Netz	45
10	Informationstechnische Konzeption und Realisierung	47
10.1	Grundlegende Betrachtungen	47
10.1.1	Aus Sicht der fachlichen Anwendung	47
10.1.2	Aus Sicht der Informationstechnik	48
10.2	Architektur und Plattform	49
10.2.1	Betriebssysteme	49
10.2.2	Programmierung	50
10.2.3	Verwendete Plattform	51
10.3	Systemkomponenten – Übersicht	53
10.4	Indexierung.....	53
10.4.1	Aufgabenstellung.....	53
10.4.2	Recherche	54
10.4.3	Probleme	54
10.4.4	Integration.....	55
10.4.5	Das Crawling	55
10.4.6	Suche	59
10.4.7	Verbindung zum Lexikon und dem semantischen Netz	61
10.4.8	Umsetzung	61
10.4.9	Resümee	61
10.5	Semantisches Netz	62
10.5.1	Verwendung in IRIS	63
10.5.2	Software-Recherche.....	63
10.5.3	Realisierung	64
10.6	Lexikon	66
10.6.1	Wiki.....	66
10.6.2	JSPWiki	66

10.7	Lernen	67
10.7.1	Einführung	67
10.7.2	Grundüberlegungen für IRIS	67
10.7.3	Realisierung	68
10.7.4	Ausblick	70
11	IRIS-Software im Überblick	71
11.1	Erläuterungen zur Verwendung von IRIS	71
11.2	Gliederung der Anwendung	72
11.2.1	Gestaltung und Aufbau der Oberfläche	72
11.2.2	Navigation	74
11.2.3	Hilfe	74
11.3	Anmeldung am System	75
11.4	Suche	76
11.4.1	Eingabe der Suche	76
11.4.2	Suchergebnis	78
11.4.3	Such- und Darstellungsoptionen	79
11.5	Redaktion	81
11.5.1	Semantisches Netz	81
11.6	Administration	83
11.6.1	Zusammenhang zwischen Benutzern, Partnern und Index	84
11.6.2	Benutzerverwaltung	84
11.6.3	Partnerverwaltung	86
11.6.4	Indexierung	87
11.6.5	Dokumentenverwaltung	95
11.6.6	Wortlistenverwaltung	98
11.6.7	Lernen	101
11.6.8	Verwaltung automatisierter Aufgaben	104
11.7	Erweiterungen von Regain	104
12	Ausblick	112
13	Glossar	113
14	Literaturverzeichnis	116
14.1	Grundlagen	116
14.1.1	Mathematik	116
14.1.2	Begriffssysteme	116
14.2	Informatik	116
14.2.1	Informationssysteme	116
14.2.2	Programmierung	117
14.2.3	Software Engineering	117
14.2.4	Persistenz	118
14.2.5	Netzwerk	118
14.2.6	Application Server	118
14.2.7	Indexierung	118
14.2.8	Semantisches Netz	119
14.2.9	Visualisierung	119
14.2.10	Lernen	120
14.2.11	Dokumentenmanagement	121
14.2.12	Abrechnungssysteme, Affiliate, Foren und Weblogs	121
14.3	Holzbau	122
14.3.1	Internet-Präsenzen	122
14.3.2	Patentwesen	122
14.3.3	Statistik	122
14.3.4	Sonstiges	122
15	Anhang	123
15.1	Technische Systeminformationen	123
15.1.1	Gesamtsystem	123
15.1.2	Unabhängige Teilanwendungen	123
15.1.3	Suche	124
15.1.4	Kommunikation zwischen Kernsystem und Lexikon	124
15.1.5	Technologien	124
15.1.6	Ordner und Dokumente	125
15.1.7	Umsetzung des MVC-Paradigmas (Model-View-Controller) ..	125
15.2	Anwendungsfälle	126
15.2.1	Akteure und Anwendungsfälle im Überblick	126
15.2.2	Anwendungsfälle der Benutzer	126

15.2.3	Anwendungsfälle der Redakteure	127
15.2.4	Anwendungsfälle des Administrators.....	127
15.3	Datenbankstruktur	128
15.4	Quellcode	128
15.4.1	Pakete	128
15.4.2	Projektstruktur	130
15.4.3	Modifizierte Software von Drittanbietern.....	132
15.4.4	Bibliotheken von Drittanbietern.....	132
15.4.5	Quellcode-Anpassungen der verwendeten Fremdprojekte	133
15.5	Build Prozess	141
16	Impressum.....	142

4 Tabellenverzeichnis

Tabelle 1: Überarbeiteter und angepasster Balkenplan der Teilprojekte 21 und 22.....	19
Tabelle 2: Auszug aus der Bestandsaufnahme und Analyse bestehender Datenbank- und Informationssystemlösungen	25
Tabelle 3: Matrix zur Definition der Hauptbenutzergruppen	32
Tabelle 4: Zusammenstellung relevanter Begriffsdefinitionen zur Fachsystematik	38
Tabelle 5: Zusammenstellung vorhandener Systematiken aus unterschiedlichen Bereichen	40
Tabelle 6: Attribute eines Dokumenten-Eintrages	96
Tabelle 7: Berechtigungsstufen für Dateizugriff (Benutzer-Level)	97
Tabelle 8: Status von Wortlisteneinträgen.....	98
Tabelle 9: Auswirkungen von Wortlisten-Status.....	98
Tabelle 10: Attribute von Begriffen und Beziehungen im semantischen Netz	101
Tabelle 11: Von Regain standardmäßig genutzte Felder in Lucene.....	106
Tabelle 12: Einstellungsmöglichkeiten zum Speicherungsverhalten von Feldern in Lucene.....	107
Tabelle 13: Einstellungsmöglichkeiten zum Indexierungsverhalten von Feldern in Lucene.....	107
Tabelle 14: Java-Pakete in IRIS	128
Tabelle 15: Quellcode Teilprojekte	130
Tabelle 16: Modifizierte Fremdsoftware	132
Tabelle 17: Verwendete Software-Bibliotheken	132
Tabelle 18: Farbcodierung zur Darstellung von Quellcode-Modifikationen.....	133
Tabelle 19: Modifikationen am Quellcode von JdbcProvider	133
Tabelle 20: Modifikationen am Quellcode von JSPWiki.....	134
Tabelle 21: Modifikationen am Quellcode von Regain	136

5 Abbildungsverzeichnis

Abbildung 1: Projektablauf, unterteilt in drei Phasen.....	19
Abbildung 2: Angepasster Projektstrukturplan. Differenziert in inhaltliche und softwaretechnische Realisierung	20
Abbildung 3: Branchen und Anteile am Gesamtumsatz von 78 Mrd. € der Holzwirtschaft	21
Abbildung 4: Gegenüberstellung der Beschäftigtenzahlen in den unterschiedlichen Branchenbereichen der deutschen Holzwirtschaft	22
Abbildung 5: Stetige Zunahme der Internetnutzer seit 1995.....	23
Abbildung 6: Zugang der Mitarbeiter zur Informationstechnologie in Abhängigkeit der Unternehmensgröße (Baden-Württemberg, Stand 2005).....	23
Abbildung 7: Gesamtübersicht Computernutzung, Internetzugang Breitbandanschluss, Firewall und Homepage in Abhängigkeit von der Unternehmensgröße (Baden-Württemberg Stand 2005)	24
Abbildung 8: Transformationsprozess – generieren von Wissen über Informationen und Daten	26
Abbildung 9: Vision einer Informationsplattform für den Holzbereich	27
Abbildung 10: Ideensammlung zum Inhalt eines übergeordneten Informationssystems.....	27
Abbildung 11: Zeitliche Gliederung des Projekts in 3 Phasen innerhalb der Projektlaufzeit. Hervorgehoben sind die inhaltlichen und technischen Arbeitsinhalte der Projekt-Phase 1	29
Abbildung 12: Inhalte und Tätigkeiten der 2. Projektphase.....	29
Abbildung 13: Inhalte und Tätigkeiten der 3. Projektphase.....	30
Abbildung 14: Benutzer der Informationsplattform und Unterscheidung in die Kategorien Profi und Non-Profi (linke Grafik). Beziehungsdreieck der Personenkreise innerhalb des Dienstleistungssystems (rechte Grafik)	32
Abbildung 15: Untergliederung der Informationsplattform IRIS in fünf Hauptbereiche sowie die Zugänglichkeiten für Benutzer, Redaktion und Administration	33
Abbildung 16: Definition der Dokumente und Zuordnung zu den definierten Qualitätsstufen	34
Abbildung 17: Hierarchie	39
Abbildung 18: Polyhierarchie, bspw. Darstellung eines Telefonnetzes (rechtes Bild) ...	39
Abbildung 19: Vernetzung, bspw. Diamantstruktur (rechtes Bild).....	39
Abbildung 20: Die Fachsystematik Holzbau im mehrdimensionalen Raum	43
Abbildung 21: Schema holzbauspezifische Systematik	44
Abbildung 22: Schematischer Ausschnitt aus dem semantischen Netz	45
Abbildung 23: Holzbauspezifische Systematik am Beispiel Holzhaus. (Quelle: Hagebau Holzhandel)	46
Abbildung 24: Prinzipieller Ablauf von Indexierung und Suche	53
Abbildung 25: Crawling-Prozess in Regain	55
Abbildung 26: Schritte beim Crawling in IRIS	57
Abbildung 27: Crawling-Schritt 1: Einlesen der Dokumente.....	58

Abbildung 28: Crawling-Schritt 2: Indexierung der Inhalte	58
Abbildung 29: Beispiel für die Suche in Regain	59
Abbildung 30: Beispiel für die Verarbeitung eines Dokuments in IRIS	60
Abbildung 31: Das allgemeine Tripel und ein Beispiel für eine konkrete Aussage.....	62
Abbildung 32: Beispiel für einen Auszug aus dem semantischen Netz	62
Abbildung 33: Struktur des semantischen Netzes	63
Abbildung 34: Die Startseite von IRIS	72
Abbildung 35: Das IRIS-Glossar als Beispiel für ein Pop-up-Fenster	73
Abbildung 36: Beispiel für Stellen zum Aufruf kontextspezifischer Hilfe.....	74
Abbildung 37: Login-Dialog.....	75
Abbildung 38: Dialog zum Anlegen eines Benutzers	75
Abbildung 39: Beispielhafte Suche nach dem Wort Holzbau.....	76
Abbildung 40: Beispiel eines Suchergebnisses	78
Abbildung 41: Beispiel für die graphische Darstellung des semantischen Netzes	78
Abbildung 42: Beispiel für die Auflistung sprachlicher Alternativen	78
Abbildung 43: Detailbeispiel beim Index-Suchergebnis.....	79
Abbildung 44: Navigation der Such- und Darstellungsoptionen	79
Abbildung 45: Beispiel für eine übermäßig umfangreiche Darstellung	79
Abbildung 46: Dialog für die Such-Optionen.....	80
Abbildung 47: Redaktionsfunktionen.....	81
Abbildung 48: Auflistung der Begriffe des semantischen Netzes.....	82
Abbildung 49: Bearbeiten eines einzelnen Eintrags.....	82
Abbildung 50: Administrationsfunktionen	83
Abbildung 51: Benutzer auflisten und bearbeiten	84
Abbildung 52: Benutzerkonto bearbeiten	85
Abbildung 53: Neuen Projektpartner anlegen.....	86
Abbildung 54: Projektpartner auflisten und bearbeiten.....	86
Abbildung 55: Übersicht der Indexierungs-Konfigurationen	87
Abbildung 56: Editieren einer noch leeren Indexierungs-Konfigurationsdatei	88
Abbildung 57: Ausschnitt aus einer Indexierungs-Konfigurations-Vorlage.....	89
Abbildung 58: Editieren einer vorhandenen Indexierungs-Konfigurationsdatei	90
Abbildung 59: Aufruf des Crawlers von der Kommandozeile	90
Abbildung 60: Quartz Scheduler Server nicht ansprechbar	92
Abbildung 61: Übersicht der Automatisierten Aufgaben.....	92
Abbildung 62: Neue Aufgabe festlegen	93
Abbildung 63: Editieren einer vorhandenen Aufgabe	94
Abbildung 64: Überblick der Dokumenten-Einträge.....	95
Abbildung 65: Details eines Dokumenten-Eintrages.....	96
Abbildung 66: Bearbeitung von Wortlisten	99
Abbildung 67: Lernen im semantischen Netz.....	102
Abbildung 68: Überblick Gesamtsystem	123
Abbildung 69: IRIS Teilanwendungen	123
Abbildung 70: Zugriffe bei der Suche	124
Abbildung 71: Kommunikation zwischen Kernsystem und Lexikon.....	124
Abbildung 72: Eingesetzte Technolgien	124
Abbildung 73: Struktur von Ordnern und Dokumenten	125
Abbildung 74: Verarbeitung einer Http-Anfrage in IRIS	125
Abbildung 75: Akteure und Anwendungsfälle im Überblick	126
Abbildung 76: Anwendungsfälle der Benutzer	126
Abbildung 77: Anwendungsfälle der Redakteure.....	127
Abbildung 78: Anwendungsfälle des Administrators	127
Abbildung 79: Übersicht Datenbank Tabellenstruktur	128
Abbildung 80: Zuordnung von Projekten zu Teilanwendungen	132
Abbildung 81: Projektabhängigkeiten	141

6 Projektbeschreibung

6.1 Ausgangslage

Innovation und Kostensenkung sind zwei wesentliche Faktoren für die Zukunftsfähigkeit im Bereich des Holz- und Holzhausbaus, sowohl der industriell strukturierten Unternehmen, als auch der kleinen und mittleren Unternehmen (KMU). Ebenso bestimmende Faktoren sind die Markt-, Produkt- und Kundenorientierung. Der Aktionsraum der Unternehmen verlagert sich auch im Holz- und Holzhausbau weg von den regionalen bzw. nationalen Märkten hin zu den internationalen Märkten.

Zur Bewältigung der hieraus resultierenden vielfältigen Aufgaben und Anforderung sind zeitnahe und verlässliche Daten und Informationen erforderlich. Das stellt die zielgerichtete Beschaffung, Bereitstellung und Verarbeitung von Informationen und Daten als einen wettbewerbsentscheidenden Faktor in den Mittelpunkt unternehmerischen Handelns. Neue wissenschaftliche Erkenntnisse, Normen oder Produktentwicklungen werden in immer kürzeren Zyklen publiziert. Demnach kann dem Informationsmanagement eine Schlüsselfunktion im Unternehmen zugesprochen werden, denn – im Vergleich zu den bisherigen klassischen Produktionsfaktoren – gewinnt die „Information“ als „neuer Produktionsfaktor“ mehr und mehr an Bedeutung, mit zunehmender Tendenz auch im Bereich der klein und mittelständisch aufgestellten Unternehmen (KMU).

Informationen und Wissen – verstanden als wissenschaftlich qualifizierte Fachinformation – gewinnt für jedermann in der Gestaltung seines privaten und beruflichen Lebens eine immer größere Bedeutung. Um dies bewerkstelligen zu können, sind effiziente und leistungsfähige Informationsbeschaffung und -management notwendig. Grundlage des Wissenstransfers ist, dass aus Daten Informationen und aus Informationen Wissen generiert werden. Hierfür benötigt man Zusammenhänge und Beziehungsmuster. Als Recherchewerkzeuge stehen zahlreiche Bibliotheken, Übersichten und Datenbanken sowohl in gedruckter als auch elektronischer Form zur Verfügung. Diesen Informationsquellen ist gemein, dass sie bei der Recherche überwiegend vorgegebene oder vorformulierte Verknüpfungen und Auswahlkriterien (Schlüsselbegriffe, Keywords) zulassen.

Die bisherige Vorgehensweise der Informationsbeschaffung und -verwaltung gestaltet sich so:

- Starre Klassifizierung nach Art eines Katalogs bzw. vorgegebene Suchbegriffe führen bei Anfragen zu umfangreichen und unübersichtlichen Ergebnisanzeigen und sind somit unbefriedigend
- Einzelne Dokumente mit Schlagwörtern versehen bspw. inhaltliche Klassifizierungen für nationale und internationale Patentschriften
- Innerhalb von schriftlichen Dokumenten kann teilweise zusätzlich nach einzelnen Wörtern gesucht werden, was aber bei umfangreichen Datenmengen zu langen Wartezeiten führt
- In der Regel bleibt die Relevanz eines Suchbegriffes im Volltext unberücksichtigt
- Querverweise und Wechselbeziehungen zu artverwandten oder tangierten Themen bleiben außer acht

Das Internet als Zukunfts-Medium zur Informationsbeschaffung bietet mit seinen gegenwärtig bekannten Suchmaschinen und Plattformen hierfür aber nur eingeschränkt eine brauchbare Basis. Aufgrund der Informationsflut findet nur eine unzureichende Bündelung und Verdichtung der fachspezifischen Informationen statt. Wesentliches Defizit ist, dass die Daten im Internet verstreut und inhaltlich nicht (logisch) miteinander vernetzt vorliegen. Dadurch bedingt ist eine Online-Recherche (z. B. Literaturrecherche) zeitaufwendig, lückenhaft und wenig komfortabel. Ferner ist es für die bisherige Art der Informationsgewinnung erforderlich, Kenntnisse in der Fachterminologie zu haben, sowie die „richtigen“ Schlagwörter (Keywords) zu verwenden.

Im Entwicklungs- und Entwurfsprozess greift man häufig auf umfangreiche Informationen zurück, welche in unterschiedlicher Form und in verschiedenen Quellen als „Insellösungen“ vorliegen, wie z. B. Datenbanken für Forschungsberichte, Patente, Produktentwicklungen, Normen und diverse Internetplattformen. Zur Deckung des hohen Informationsbedarfs – gerade in den KMU – stehen überwiegend bescheidene personelle und infrastrukturelle Ressourcen für ein professionelles Informationsmanagement zur Verfügung. Ferner ist dem erkennbaren Trend weg von gedruckten hin zu elektronischen Medien in geeigneter Weise zu begegnen.

Speziell im Holzbereich, für kleine und mittelständische Unternehmen (KMU) sowie für Hochschulen und Institute besteht also ein dringender Bedarf für ein Informationssystem, das folgende Voraussetzungen berücksichtigen muss:

- Innovationen, Kostensenkung und Wettbewerb (Internationalisierung) dominieren
- Zunahme der Produktkomplexität
- Stetige Zunahme der Bedeutung des Produktions-, Wirtschafts- und Erfolgsfaktors „Information“, bei gleichzeitigem Rückgang der bisherigen klassischen Produktionsfaktoren
- Rasante quantitative Zunahme an Publikationen und bereitgestellten Informationen, z. B. über das Medium „Internet“, ohne qualitative oder fachliche Zuordnung oder erkennbare Beziehungsmuster
- Fehlende organisatorische Strukturen und Kenntnisse in der Beschaffung, Nutzung und Weiterverarbeitung von Informationen
- Begrenzte personelle Ressourcen und Kenntnisse der KMU im IT-Bereich und in der Informationsbeschaffung
- Informationsvielfalt der Quellen ist für den Benutzer mit Intransparenz und Heterogenität verbunden, d. h. wachsender Aufwand, erhöhter Zeitbedarf und Überforderung des Suchenden
- Trend weg von gedruckten Medien hin zu elektronischen Medien

6.2 Projektziel

Für die Informationsbeschaffung zu forschungsrelevanten Themen oder für die Entwicklung neuer Produkte und Technologien stehen dem Anwender im Bauwesen allgemein, so auch im Holzbau, gegenwärtig eine Vielzahl von „unkoordinierten Insellösungen“ zur Verfügung. Ein übergeordnetes Dienstleistungssystem, welches für Forschungs- und Entwicklungsprozesse im Holzbau den Abruf strukturierter Daten und Informationen in digitaler Form ermöglicht, fehlt bislang. Das Projektziel wird wie folgt formuliert:

„Es ist ein übergeordnetes Dienstleistungssystem zu entwickeln, das für Forschungs- und Entwicklungsprozesse im Holzbau den Abruf strukturierter Daten und Informationen ermöglicht. Des Weiteren dient das Dienstleistungssystem als Schnittstelle zu vorhandenen Daten- und Informationsbeständen.“

Ferner sind mit IRIS

„... Forschungs- und Entwicklungsarbeiten... zielgerichtet zu unterstützen und damit effektiver zu gestalten...“

IRIS¹ – konzipiert als ein sich selbst entwickelndes, „lernendes“ internetbasiertes Informationssystem – soll Querverbindungen und interdisziplinäre Verknüpfungen hinsichtlich der Entwicklungsprozesse im Holzbau aufzeigen. Die Verfolgung der Verknüpfungsmöglichkeiten soll durch eine neuartige, kompakte und dennoch übersichtliche Darstellung der Suchergebnisse erleichtert werden. Ein mehrdimensionales, nichtlineares Navigieren unterstützt dabei den Benutzer und bietet eine einfache und intuitive Bedienbarkeit. Ähnliche Ansätze sind als "Star-Browsing" und visualisierte, hierarchische Baumstrukturen bekannt.

Die Ausgangslage der KMUs erfordert ein branchenweites Informationssystem. Die Notwendigkeit eines flexiblen, an den Erfordernissen der Praxis orientierten Informationssystems, mit einer verbesserten Transparenz des Informationsangebots, ist hier gegeben. Forschung und Lehre profitieren in gleicher Weise von einer zentralen, branchenbezogenen Informationsplattform.

¹ IRIS – Integriertes Relationales Informationssystem

Anders als ursprünglich geplant, erfolgt die Projektbearbeitung mit TUM- und HR-Mitarbeitern ausschließlich an der HR. Durch die räumliche Zusammenlegung können Synergie-Effekte effizient genutzt und der Aufwand für Koordination und Kommunikation minimiert werden. Der Personaleinsatz teilt sich 40/60 auf für Inhalt und – mit leicht verstärktem Fokus – für die Softwareentwicklung. Diese Differenzierung kann dem angepassten Projektstrukturplan entnommen (Abbildung 2) werden.

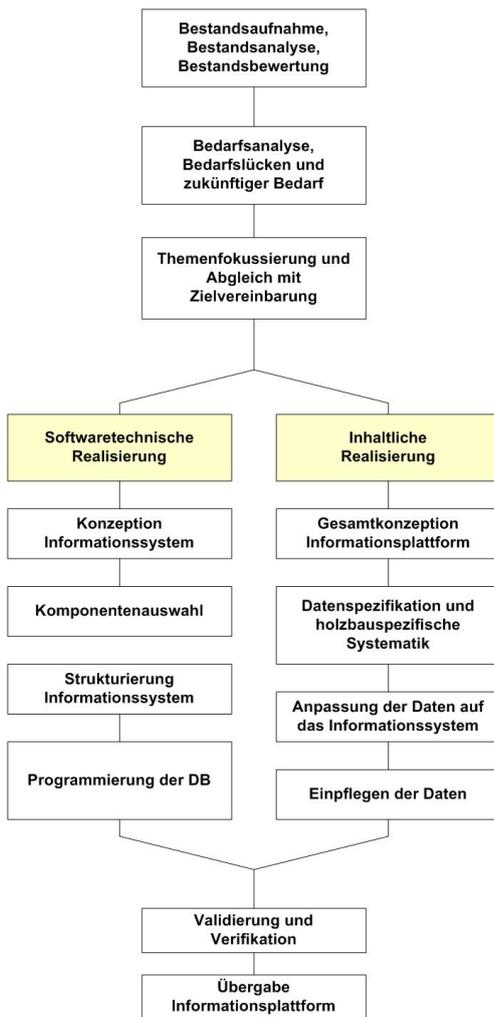


Abbildung 2: Angepasster Projektstrukturplan. Differenziert in inhaltliche und softwaretechnische Realisierung

6.4 Abgleich mit dem Projektantrag

Die im Projektantrag formulierten Arbeitspunkte und Inhalte wurden dem Stand der Technik angepasst. Dies hängt mit der technischen Weiterentwicklung im Bereich der Informations- und Kommunikationstechnologie zusammen.

- Datenträger² für lokale Recherche werden nicht mehr benötigt
- Keine CAD- und Objektdatenübernahme aus IRIS
- Bildanalyse wird aus dem Aufgabenpaket herausgenommen
- Digitalisierung von Printmedien wird nur exemplarisch durchgeführt

² Prof. Kurt Schwaner, FH Biberach/Riß: Gespräch zum Thema „Holzbank“ und Datenträger-Problematik

7 Bestandsaufnahme

Es wurden Informationen und Fakten zur Holzbranche und zum IT-Verständnis recherchiert sowie bestehende Daten- und Informationssysteme hinsichtlich ihrer Inhalte und Leistungsfähigkeit analysiert und bewertet. Daraus folgende Erkenntnisse flossen in die Konzeption der branchenweiten Informationsplattform ein.

7.1 Beschreibung des Branchenumfelds – Strukturdaten

Die dezentrale Branchenstruktur im Holzbereich spiegelt sich auch im Bereich des Holz- und Holzhausbaus wieder. Studien³ zeigen, dass ca. 65% der Unternehmen weniger als 10 und etwa 30% sogar weniger als 5 Beschäftigte haben. So gibt es z. B. in Bayern über 4.000 holzverarbeitende Unternehmen, von denen die meisten weniger als 10 Mitarbeiter haben. Auf Basis dieser Zahlen kann man davon ausgehen, dass hier im IT-Bereich und Informationsmanagement keine bzw. nur geringe Kenntnisse und personellen Ressourcen vorhanden sind.

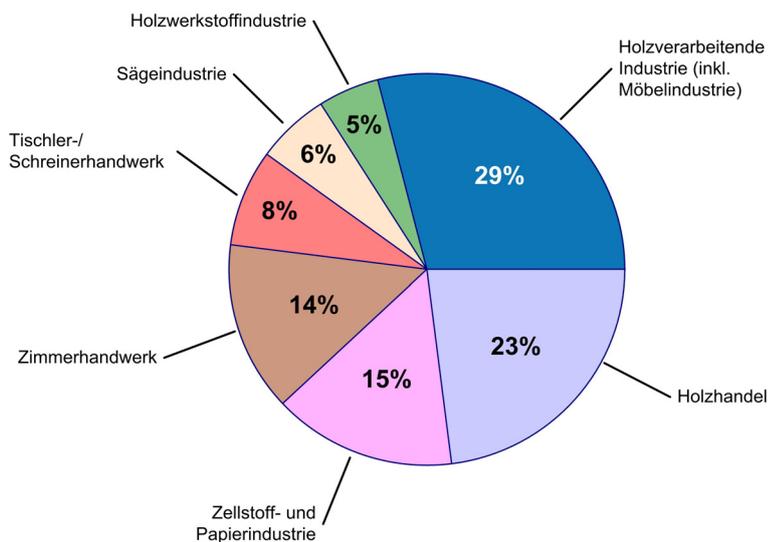


Abbildung 3: Branchen und Anteile am Gesamtumsatz⁴ von 78 Mrd. € der Holzwirtschaft

³ Zimmerer 2010: Studie Modul 1 Mitgliederfragung, Roland Berger, Market Research

⁴ Quelle, Deutscher Holzwirtschaftsrat, 2004, Abdruck in „Natürlich Holz“, Die deutsche Holzwirtschaft Zahlen und Fakten, Holzabsatzfonds. Bonn, 2005

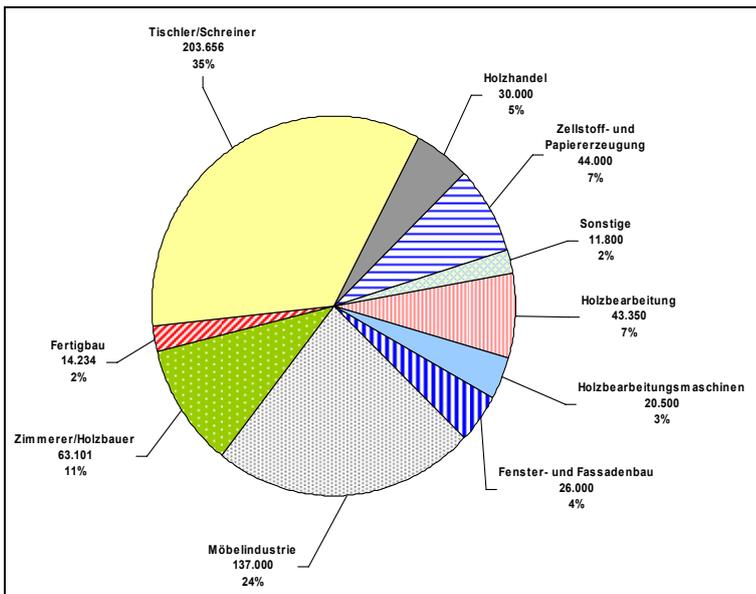


Abbildung 4: Gegenüberstellung der Beschäftigtenzahlen in den unterschiedlichen Branchenbereichen der deutschen Holzwirtschaft⁵

7.2 IuK – Zugangsmöglichkeiten

Neben der Branchenstruktur sind für die Konzeption einer internetbasierten Informationsplattform Kenntnisse über die Zugangsmöglichkeiten der betroffenen Personenkreise zu Informations- und Kommunikations-Technologien (IuK) relevant. Entsprechende Recherchen sind erfolgt und Vergleiche⁶ sind angestellt worden.

Die Zahl der Internetnutzer ist in Deutschland seit 1997 von ca. 5 Mio. auf nun über 30 Mio. angewachsen. Für die Holzbranche liegen keine zuverlässigen Zahlen vor, man kann annehmen, dass ein ähnlicher Trend vorliegt. Das Internet wird zunehmend ein nicht substituierbarer Bestandteil im täglichen Geschäfts- und Arbeitsablauf. In Deutschland haben 94% aller Unternehmen mit mehr als 10 Mitarbeitern einen Internetzugang.⁷ Neben dem Zugang zu Computer und Internet ist die Art dieses Zugangs zu betrachten. 44% der Unternehmen stellen den Zugang zum Internet über eine ISDN-Verbindung her. Der Trend geht zu Breitbandverbindungen wie z. B. DSL oder ADSL. Mittlerweile verfügen 44% der Unternehmen, die einen Internetzugang nutzen, über einen Breitbandzugang⁸.

⁵ Quelle, Natürlich Holz, Die deutsche Holzwirtschaft Zahlen und Fakten, Holzabsatzfonds. Bonn, 2005

⁶ Neben Statistiken und Zahlenmaterial des Statistischen Bundesamts zu Unternehmens-Erhebungen im Bereich IuK wird auf Datenmaterial des Statistischen Landesamts Baden Württemberg Bezug genommen. Nach Informationen⁶ des Statistischen Landesamts Bayern sind speziell für Bayern hierzu keine Daten erhoben worden. Baden-Württemberg kann aufgrund ähnlicher Rahmen- und Strukturbedingungen für eine qualitative Aussage und Betrachtung der Thematik „Unternehmen - Computer - Internet“, für Trends und Tendenzen herangezogen werden. Dieses Vorgehen ist nach Rücksprache mit dem Statistischen Landesamts Bayern für einen orientierenden Vergleich als unproblematisch zu sehen.

⁷ [SBaPK06], Statistisches Bundesamt „Aktuelle Nutzung von Informationstechnologie in Deutschland“ am 21.02.06, Frankfurt/Main

⁸ [IuKBW05] „Informations- und Kommunikationstechnologie in Unternehmen 2005“, Statistisches Monatsheft Baden-Württemberg 3/2006, Dr. Richard Kössler

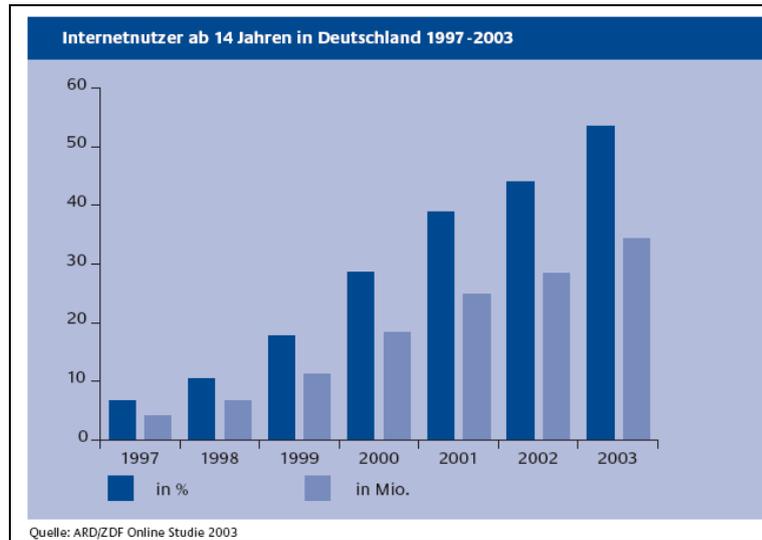


Abbildung 5: Stetige Zunahme der Internetnutzer seit 1995⁹

Die IuK-Durchdringung hängt von der Unternehmensgröße und -tätigkeit (Branche) ab.¹⁰ Zum Vergleich haben Unternehmen, die sich mit Forschung und Entwicklung (FuE) befassen, eine wesentlich höhere IuK-Durchdringung als Unternehmen des Baugewerbes. Das liegt darin begründet, dass im Bereich von FuE der Einsatz und die Nutzung von Computer und Internet unumgänglich sind, im Baugewerbe dagegen viele Unternehmen für die Erledigung ihrer Aufgaben nicht ausnahmslos auf den Einsatz von IuK-Technologien angewiesen sind.

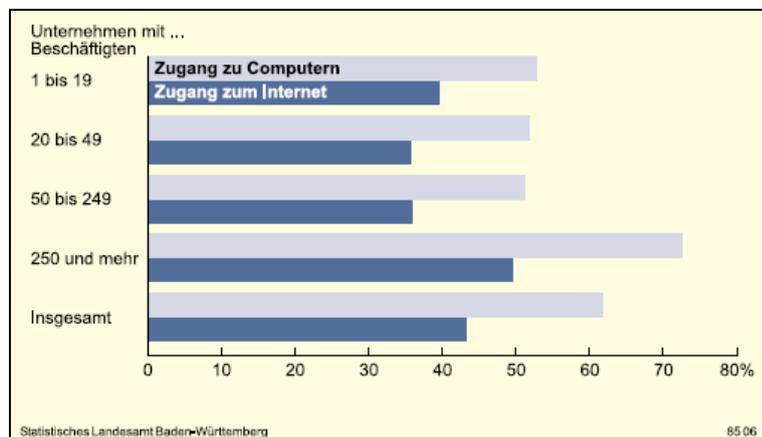


Abbildung 6: Zugang der Mitarbeiter zur Informationstechnologie in Abhängigkeit der Unternehmensgröße (Baden-Württemberg, Stand 2005)⁹

⁹ [IuKBW05], „Informations- und Kommunikationstechnologie in Unternehmen 2005“, Statistisches Monatsheft Baden-Württemberg 3/2006, Dr. Richard Kössler

¹⁰ [IuKBW05] „Informations- und Kommunikationstechnologie in Unternehmen 2005“, Statistisches Monatsheft Baden-Württemberg 3/2006, Dr. Richard Kössler, Seite 21

Zusammengefasst kann festgestellt werden, dass IuK-Technologien und das Internet in der Zukunft noch weiter in die Unternehmen vordringen. Geschäftsabläufe werden darauf ausgerichtet. Der Trend geht zu schnelleren Breitband-Internetverbindungen, um alle Möglichkeiten des Internets zu nutzen. Je nach Wirtschaftszweig ist der Einsatz von IuK-Technologie mehr oder weniger erforderlich. Die Notwendigkeit und die Nutzung von IuK-Technologie und Internetzugang ist in den Bereichen Forschung und Entwicklung als ein wesentliches Kriterium zu sehen. Gewiss wird es auch bei den KMUs des Baubereichs zu einer weiteren Durchdringung und zur Ausweitung von IuK-Technologie kommen. Ein Grund dafür ist auch, dass derzeit Behörden auf Online-Transaktionen umstellen (z. B. Finanzbehörden).

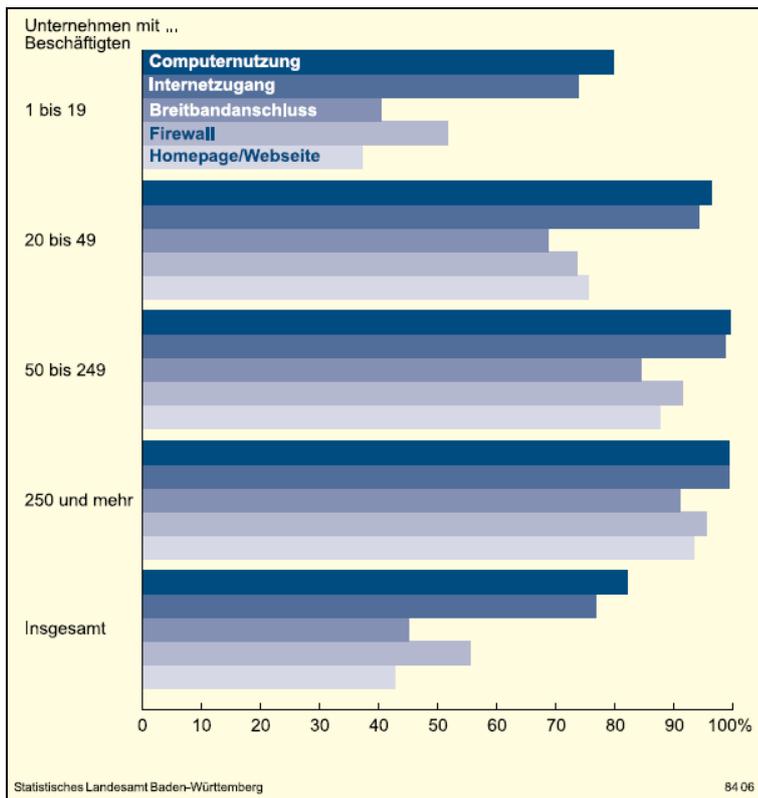


Abbildung 7: Gesamtübersicht Computernutzung, Internetzugang Breitbandanschluss, Firewall und Homepage in Abhängigkeit von der Unternehmensgröße (Baden-Württemberg Stand 2005)⁹

7.3 Bestandsaufnahme und Analyse bestehender Datenbank- und Informationssystemlösungen

Neben allgemeinen Themen wurden auch Informationsplattformen und Datenbanken hinsichtlich deren

- Themengebiete bzw. Themenbereiche,
- Inhalte (Tabellen, Dokumente etc.),
- Bedienung und Benutzerführung
- und inhaltlicher Strukturierung (Systematik, Klassifizierung)

näher betrachtet und untersucht.

Die Integration mehrerer Systembestandteile, die in IRIS angestrebt wird, konnte bei keinem der untersuchten Systeme vorgefunden werden.

Tabelle 2: Auszug aus der Bestandsaufnahme und Analyse bestehender Datenbank- und Informationssystemlösungen

Kriterium	Inhalt	Aktualität	Datenausgabe	Downloadbereich	Benutzerfreundlichkeit	*Anmerkung
Untersuchte Datenbank	k. A.	++	0	-	0	
IHD Holzbank	0	-	-	-	+	Literatursuche - nur Metadaten
Bauteildaten.de	k. A.	+	-	?	++	Literatursuche - nur Metadaten
RSWB	k. A.	+	-	?	++	Literatursuche - nur Metadaten
Projektdatenbank DGfH	k. A.	++	0	-	++	Literatursuche - nur Metadaten
Baustoffchemie.de	0/++	++	+	k. A.	+	Niveau und Autoren unterschiedlich
Datenbank Werkstoffe TUM	+	-	-	-	++	
Eurolognum Datenbank	k. A.	k. A.	k. A.	k. A.	++	Verlinkung zu Mitgliedern/Partner
Fraunhofer Holzkirchen	+	++	++	++	+	
ELFIS Forst- Holzwirtschaft BFH	k. A.	0	-	-	++	Literatursuche - nur Metadaten
BFH LIT	k. A.					Literatursuche - nur Metadaten
Dataholz.com	+	++	+	+	++	
Wikipedia	0	++	0	+	++	

Bewertung	++	+	0	-	k. A.
	sehr gut	gut	Standard	nicht vorhanden	Keine Angaben

8 Inhaltliche Konzeption und Realisierung

Die in Projektphase 1 vorgenommene inhaltliche Konzeption des übergeordneten Dienstleistungssystems (Informationsplattform) wird nachfolgend zusammengefasst dargestellt, ohne dass auf alle Punkte im Detail eingegangen wird.

8.1 Motivation

Mit dem Informationssystem IRIS soll ein Beitrag zur Informationsvermittlung und Wissensteilung für alle Beteiligten im Holz- und Holzhausbau geleistet werden. IRIS soll die Benutzer individuell dabei unterstützen, aus Daten Informationen und aus Informationen Wissen zu generieren. Im Holzbereich existieren zwar diverse Internet-Plattformen und Online-Datenbanken, doch sind diese vielen potentiellen Nutzern unbekannt. Entweder haben sie keine Kenntnis von deren Existenz, das Angebot ist zu undurchsichtig oder der Zugang bleibt bestimmten Benutzern vorbehalten. Eine branchenübergreifende Informations- und Wissensteilung findet somit nur bedingt und unvollständig statt. Mit IRIS wird die Schnittstelle zu den relevanten Quellen hergestellt und gepflegt.

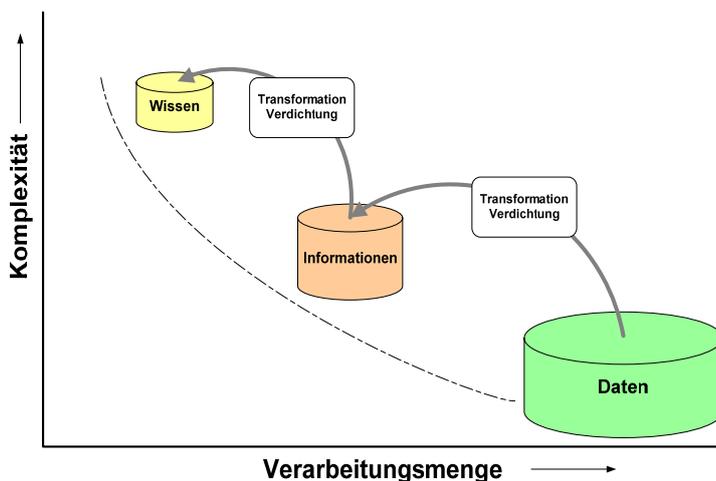


Abbildung 8: Transformationsprozess – generieren von Wissen über Informationen und Daten

Mitarbeiter verbringen ca. 25 bis 50 Prozent ihrer Arbeitszeit allein mit Informationssuche.¹¹ Dies ist nicht nur für KMUs, sondern auch für Forschungsstellen ein erheblicher finanzieller und zeitlicher Faktor. Da ca. 25% der Recherchen erfolglos bleiben, werden häufig neue Daten oder Informationen erstellt, die redundant und/oder inkonsistent sein können. Falsche und unübersichtliche Datenbestände können die Folge sein. Durch Einsatz von Informations- und Wissensmanagement ist eine Kosten- und Zeitreduzierung von annähernd 50% zu beobachten. Hier gibt es unerschlossene Potentiale für den Holzbau.¹²

¹¹ Quelle, www.wissensmanagement.net, Abruf vom 03.04.2006, Bezug zur IDC-Studie

¹² Mertins, Heisig, Vorbeck, Fraunhofer IPK, 2001 (entnommen aus Martin Seidel, Die Bereitschaft zur Wissensteilung, Deutscher Universitätsverlag, 2003)

8.2 Ideensammlung zu IRIS

Der Einstieg in die Konzeption erfolgte mit mehreren Brainstormings (u. a. mit Hilfe von Mind Maps) innerhalb des Projektteams und in Gesprächen¹³ mit Projektpartnern, Unternehmen und Verbänden, um Fragen zu klären und Bedürfnisse an ein Informationssystem festzuhalten. Die erarbeitenden Kernaussagen zu Visionen und Anforderungen an IRIS sind in Abbildung 9 und Abbildung 10 dargestellt.

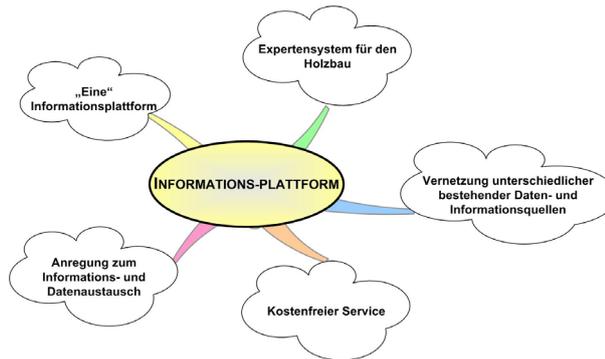


Abbildung 9: Vision einer Informationsplattform für den Holzbereich



Abbildung 10: Ideensammlung zum Inhalt eines übergeordneten Informationssystems

Aus den hergestellten Kontakten hat sich herauskristallisiert, dass es um ein komplexes Projekt geht. Die Komplexität resultiert aus dem Holz- und Holzhausbereich selbst, aber auch aus den vielen Schnittstellen zu Projektpartnern und Benutzern des Systems. Wichtigstes Ergebnis aller Gespräche war, dass die Realisierung einer Plattform für den Informationsaustausch in der Tat notwendig ist.

Das Dienstleistungssystem IRIS soll keine vorhandene Informationsplattform oder Datenbank ersetzen, sondern vor allem eine Schnittstelle zwischen den einzelnen „Insellösungen“ bieten. Die Vernetzung dieser Insellösungen wird durch eine geeignete Software und über neue Methoden realisiert. Sowohl die Information gebenden Kooperationspartner als auch die Systembetreiber profitieren von IRIS.

¹³ Gespräche mit Herrn Dr. Jauch, Publizist und ehemaliger Redakteur im DRW-Verlag, am 23.02.2006 in Rosenheim und Herrn Lange, BDF Bundesverband Deutscher Fertigbauer, am 13.03.2006 in Rosenheim sowie an mehreren Terminen mit dem ift Rosenheim, Ulrich Sieberath, Prof. Dr. B. Feindor, DGfH e.V., Egle Engineering, Fa. Lignotrend Herr Eckert und Steinbeis Forschungsstiftung Dr. Bierter

Die inhaltliche Gestaltung von IRIS erforderte die Entwicklung eines gesonderten Konzepts, des sog. Kooperations-Konzepts. (siehe Kapitel 8.5).

8.3 Projektgestaltung

Für die Projektumsetzung, innerhalb der Projektlaufzeit von 32 Monaten, wurde das Gesamtprojekt IRIS in drei mehrstufige, aufeinander aufbauende Phasen unterteilt (Abbildung 11, Abbildung 12, Abbildung 13). Die über das Projektende hinausgehende Phase „ff“ verdeutlicht, dass IRIS längerfristig weiter entwickelt werden soll.

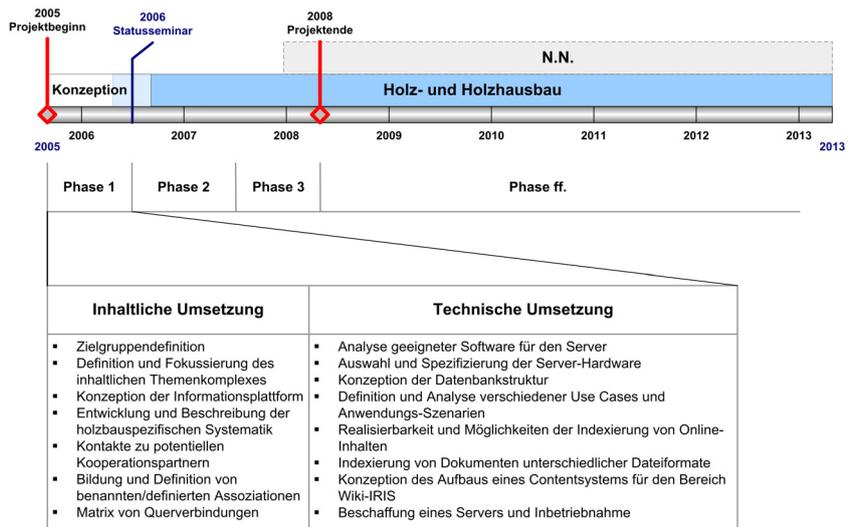


Abbildung 11: Zeitliche Gliederung des Projekts in 3 Phasen innerhalb der Projektlaufzeit. Hervorgehoben sind die inhaltlichen und technischen Arbeitsinhalte der Projekt-Phase 1

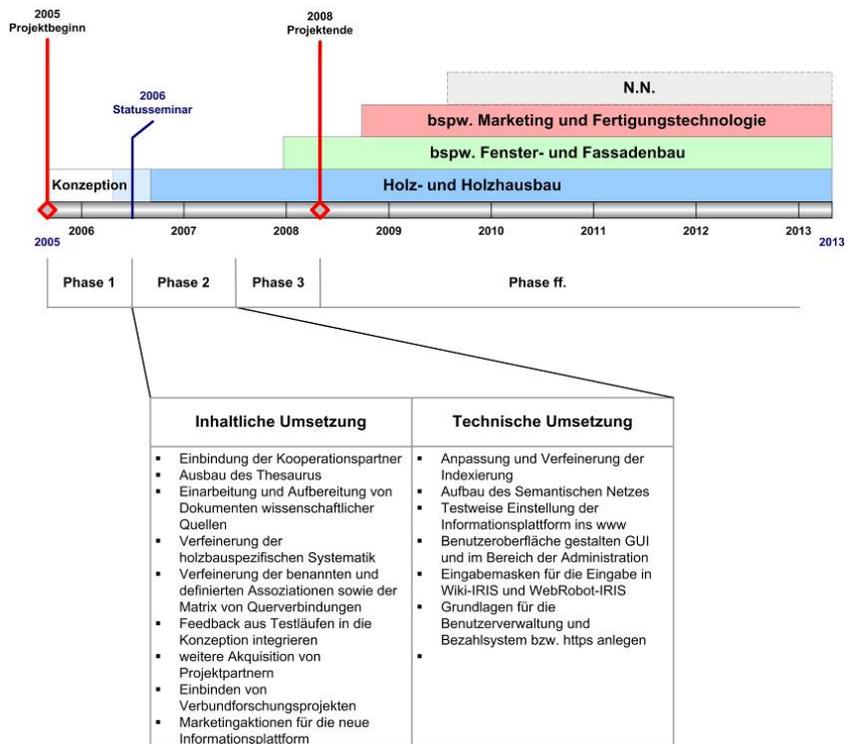


Abbildung 12: Inhalte und Tätigkeiten der 2. Projektphase

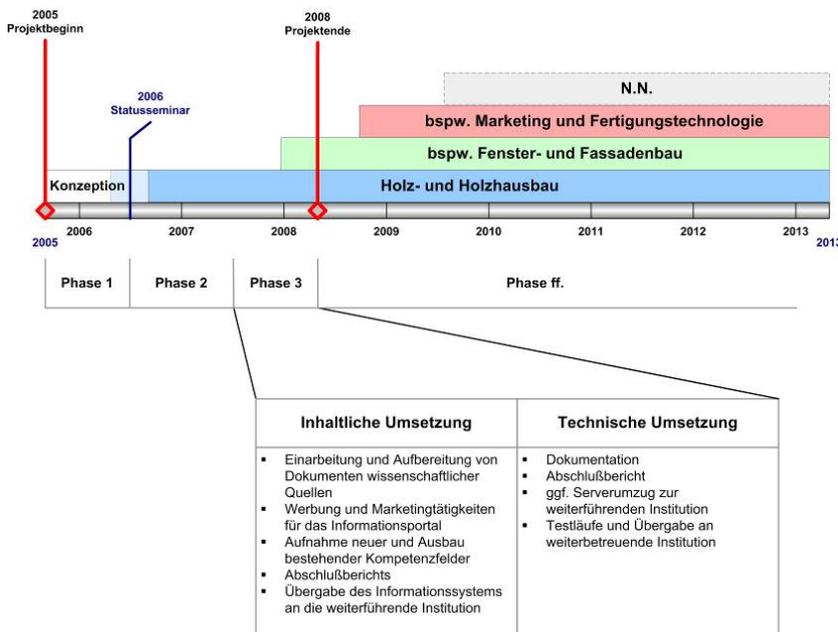


Abbildung 13: Inhalte und Tätigkeiten der 3. Projektphase

8.4 Alleinstellungsmerkmale von IRIS

Der wesentliche Punkt, der das IRIS-System von anderen Plattformen und Datenbanken unterscheidet, ist die Integration unterschiedlicher Informationssysteme wie Suchmaschine, Lexikon und semantisches Netz. Dem Benutzer von IRIS werden nur Inhalte von Quellen vermittelt, die einem definierten wissenschaftlich-technischen Anspruch entsprechen. Dies kann entweder durch die Dokumentauswahl oder durch die Quellenauswahl beeinflusst werden. Der Benutzer hat die Einflussmöglichkeit, sich bei der Recherche auf bestimmte, selektierbare Quellen, wie z. B. Hochschulen oder Institute, zu beschränken. Der Benutzer kann somit individuell, je nach Fachwissen, Bedarf oder Anforderungen, recherchieren.

Neben dem Holzbau werden auch angrenzende Themen integriert. Dies bietet für den Benutzer den Vorteil, dass er bei der Recherche nicht in unterschiedlichen Fachdatenbanken und Informationsplattformen suchen muss, sondern innerhalb einer Plattform zu mehreren Themen fachliche Informationen erhält.

Ein wesentliches Qualitäts- und Alleinstellungsmerkmal von IRIS ist die Möglichkeit der Quellenvorauswahl, die bei bekannten Internet-Suchmaschinen gerade für weniger erfahrene Nutzer nicht oder nur schwer realisierbar ist.

8.5 Vorstellung des Kooperationspartner-Konzepts

Der Grundgedanke eines Kooperationspartner-Konzepts (Wissenspool-Modell) stammt aus den ersten Treffen und orientierenden Gesprächen mit potentiellen Kooperationspartnern¹⁴ in Projektphase 1. Diese wurden persönlich angeschrieben und über das IRIS-Projekt informiert. Im nächsten Schritt kam es zu gezielten Direktgesprächen, um Meinungen, Wünsche und Probleme in Bezug auf ein übergeordnetes Dienstleistungssystem zu erörtern.

¹⁴ Projekt „MediaTUM“ der Bibliothek der TUM; Deutsche Gesellschaft für Holzforschung (als Projektpartner involviert); Institut für Fenstertechnik ift Rosenheim; Bund Deutscher Fertigbauer (BDF); Kurt Schwaner, Fachhochschule Biberach; Uwe Germeroth, HSB Biel; Dr. Jauch, DRW-Verlag

Eine Einrichtung kann nicht alle Themengebiete im Holzbau fachlich und personell abdecken. Um die Verantwortung für die Qualität und den wissenschaftlichen Anspruch der Daten und Informationen sowie den finanziellen und zeitlichen Aufwand auf mehrere Schultern zu verteilen, erscheint das Kooperationspartner-Konzept sinnvoll.

Dieses Konzept sieht die Einbindung von Experten, Hochschulen und Instituten als gleichberechtigte Partner vor. Jeder Kooperationspartner ist Pate für ein Themengebiet und steht für dessen inhaltliche Ausgestaltung und die redaktionelle Unterstützung zur Verfügung. Dies bedeutet, dass die Kooperationspartner inhaltliche Beiträge liefern, die dann in IRIS entsprechend aufbereitet und dem Benutzer zugänglich gemacht werden.

Die Einbindung von Kooperationspartnern mit deren Fach- und Kernkompetenzen bringt wesentliche Vorteile für beide Seiten. Für die Informationsplattform stehen aktuelle, qualitativ hochwertige und inhaltlich-wissenschaftlich fundierte Daten sowie Informationen zur Verfügung. Ferner profitieren durch die kontinuierliche Zusammenarbeit mit den Kooperationspartnern die Erstellung und Verfeinerung der holzbauspezifischen Systematik, die korrekte Zuordnung von Dokumenten zu den Themenbereichen und die Systempflege. Das Zusammenführen unterschiedlicher externer Quellen und Experten führt zur Ausbildung eines Netzwerks und zur Verknüpfung von „Insellösungen“. Gerade darin ist ein Vorteil für die Kooperationspartner zu sehen, die durch eine derartige Vernetzung in ihrer täglichen Arbeit unterstützt werden. Auf der anderen Seite führt die Mitwirkung an der Informationsplattform zu einem größeren Bekanntheitsgrad und somit zu einem Marktvorteil mit Unterstreichung der Fachkompetenz.

Die DGfH ist der wichtigste Kooperationspartner. Neben der inhaltlichen Unterstützung mit Forschungsberichten und Dokumenten ist sie für die Weiterführung der Informationsplattform nach Projektende zuständig. Mit der Einbindung der DGfH-Datenbank wird ein wichtiger Grundstein der wissenschaftlich-technischen Ausrichtung gelegt. Ferner soll die DGfH als Vorreiter mit Signalwirkung weitere potentielle Kooperationspartner zur Beteiligung an dem Konzept eines übergeordneten Informationssystems für den Holzbereich motivieren.

8.6 Inhaltliche Themengebiete

Inhaltlich erfolgt die Fokussierung auf den Holz- und Holzhausbau, nachfolgend Holzbau genannt. Das bedeutet, dass hauptsächlich Themenbereiche, die in (unmittelbarem) Zusammenhang mit dem Holzbau stehen, betrachtet werden. Details zum Inhalt von IRIS lassen sich der weiter unten erläuterten Fachsystematik entnehmen. Aus technischer Sicht ist die Informationsplattform so konzipiert, dass zukünftig noch weitere Themengebiete integriert werden können. Das System ist damit flexibel für zukünftige, heute noch nicht absehbare Veränderungen.

8.7 Zielgruppen und Anwender

Die Konzeption berücksichtigt die unterschiedlichen Interessengruppen von IRIS (Anwender, Akteur, Projektpartner und Systembetreiber). Anwender und Akteur können in der Rolle „Benutzer“ zusammengefasst werden (Abbildung 14 und Tabelle 3). Es lassen sich verschiedene Interaktionen im IRIS-Informationssystem definieren. Das sind Mensch-Maschine- und Maschine-Maschine Beziehungen bzw. -Interaktionen (Details hierzu siehe Lastenheft).

Generell ist eine Registrierung der Benutzer des Informationssystems erforderlich. Nur registrierten Benutzern wird die volle Funktionalität der Informationsplattform eingeräumt. Der Hintergrund ist, dass das Benutzerverhalten und die Suchvorgänge/-anfragen mit in das lernende System integriert werden. Aussagekräftige Rückschlüsse zu häufig gesuchten Begriffen und Themenbereichen sind möglich, Branchen-Trends können abgeleitet werden. Diese Informationen können wiederum für Marketingzwecke innerhalb der Holzbranche verwendet werden.

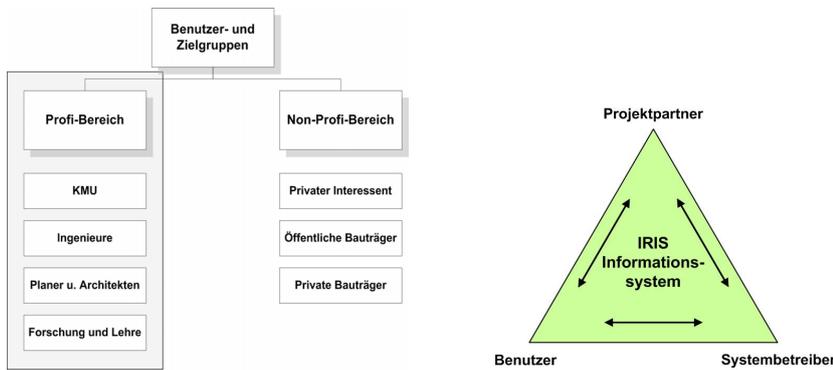


Abbildung 14: Benutzer der Informationsplattform und Unterscheidung in die Kategorien Profi und Non-Profi (linke Grafik). Beziehungsdreieck der Personenkreise innerhalb des Dienstleistungssystems (rechte Grafik)

Tabelle 3: Matrix zur Definition der Hauptbenutzergruppen

Benutzergruppe \ Themengebiet	Unternehmen Holzwirtschaft	Unternehmen Zulieferer	Ingenieure Techniker Meister	Planer Architekt	Hochschulen Institute Forschung	Berufliche- und weiterführende Schulen	Öffentliche Bauherren	Private Bauherren
Bautechnik	69%	54%	82%	62%	82%	57%	49%	37%
Fertigung	80%	75%	65%	20%	60%	45%	20%	20%
Betriebswirtschaft und Marketing	93%	90%	60%	50%	63%	40%	33%	27%
Weiterbildungs- und Informationsmöglichkeiten	63%	57%	63%	54%	74%	49%	31%	43%
Mittelwert	74%	64%	70%	52%	73%	50%	38%	34%

Prozent	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
---------	------	-------	-------	-------	-------	-------	-------	-------	-------	--------

8.8 Lastenheft und Pflichtenheft

Das Lastenheft¹⁵ für IRIS wurde auf Basis des Projektantrags, der Brainstormings und der Recherche ausgearbeitet. Das Pflichtenheft¹⁶ wurde daraus abgeleitet. Es bildet die Basis für das gesamte Software-Engineering. Kernpunkte des Lastenhefts bzw. des Pflichtenhefts sind

- Bedienung interdisziplinärer Fragestellungen
- Automatisierte Generierung und Strukturierung von Daten/Informationen
- Innovative Werkzeuge zur Suche, Navigation (semantisches Netz), Auswertung und Verknüpfung
- Möglichst keine Pflege von Metadaten
- Internes Lexikon für die Fachterminologie
- Selbstlernendes System, Aufzeigen von Wechselbeziehungen

8.9 Aufbau der Informationsplattform

Der Aufbau der Informationsplattform erfolgt unter dem Gesichtspunkt, dem Benutzer Daten und Informationen einfach zugänglich zu machen und entsprechend darzustellen. IRIS dient primär zur Literaturrecherche für Produktentwicklungen, Projektideen und Forschungsanträge. Es ist aber ebenso der Einsatz im Alltagsgeschäft angedacht. Dort ist die Informationsbeschaffung, Recherche nach Literatur, Definitionen oder Kennwerten für Konstruktion, Bemessung und Auslegung von Bauvorhaben zu sehen. Neben dem Abruf von Daten und Informationen soll auch die aktive Beteiligung an der Informationsplattform ermöglicht werden. Benutzer können in einem definierten Bereich von IRIS Begriffsdefinitionen, Artikel oder Querverweise einarbeiten, um so selbst am stetigen Ausbau der Plattform mitzuwirken.

Die Informationsplattform kann in fünf Hauptbereiche (Abbildung 15) unterteilt werden. Dem Benutzer sind dabei jedoch nur drei der fünf Bereiche zugänglich, die beiden anderen sind der Systembetreuung vorbehalten.

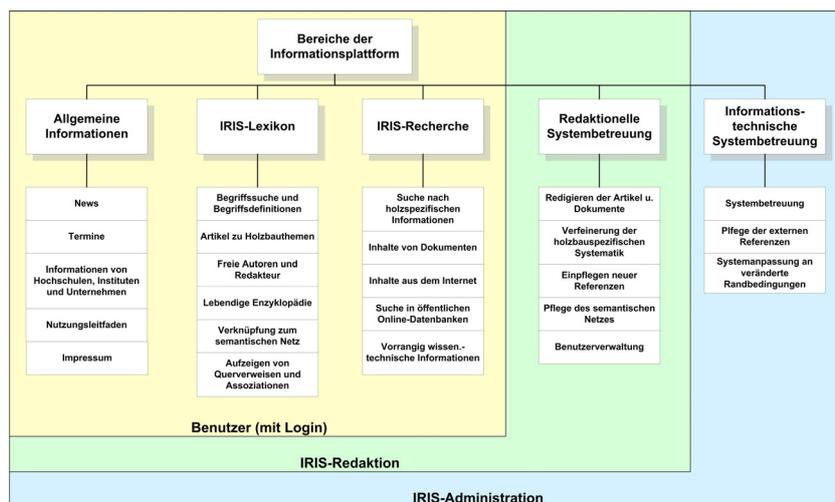


Abbildung 15: Untergliederung der Informationsplattform IRIS in fünf Hauptbereiche sowie die Zugänglichkeiten für Benutzer, Redaktion und Administration

¹⁵ Lastenheft „Gesamtheit der Anforderungen des Auftraggebers an die Lieferungen und Leistungen eines Auftragnehmers“, DIN 69905

¹⁶ Pflichtenheft „vom Auftragnehmer erarbeitete Realisierungsvorgaben aufgrund der Umsetzung des Lastenhefts“, DIN 69905

8.10 Daten, Dokumente und Artikel

Daten- und Informationsquellen für das IRIS-System können Datenbanken, Webseiten und zur Verfügung gestellte digitale Dokumente oder Metadaten im XML-Format sein. Welche Quellen und Dokumente in die Informationsplattform integriert bzw. miteinander vernetzt werden, wird durch die IRIS-Redaktion gesteuert.

Ein inhaltlicher Bestandteil sind die Dokumente. Das sind digital vorliegende Texte, Veröffentlichungen und Printmedien, die mit der Thematik „Holz und Holzbau“ in Verbindung stehen. Möglichkeiten der Dokumentanalyse gibt es für die Dateitypen Microsoft Office (doc, xls, ppt), Adobe PDF (pdf), ASCII-Text (txt), XML (nur rein textlicher Inhalt) sowie die OpenOffice-Formate (verschiedene Endungen). Zu Beginn des Projektes bildeten die Informationsdienst-Holz-Schriftreihen, Mikado-Fachartikel, „Holzbank“- und Hochschulskripte die Grundlage der inhaltlichen Erschließung. Im weiteren Verlauf wurden DGfH-Forschungsberichte und Dokumente aus dem laufenden Verbundforschungsprojekt eingebunden.

Alle Einträge im IRIS-Lexikon werden als Artikel bezeichnet. Die Unterscheidung in Artikel und Dokumente dient zur sprachlichen Abgrenzung, wenn man Aussagen über das IRIS-Lexikon und die IRIS-Recherche treffen möchte. Artikel werden im redaktionellen Bereich durch den Redakteur erstellt und sind jederzeit editierbar. Dokumente sind unveränderliche Dateien, die nicht editierbar sind bzw. nicht editiert werden dürfen, d. h. den Ursprungszustand behalten müssen, beispielsweise Informationsdienst Holz-Schriftreihen oder Dissertationen.

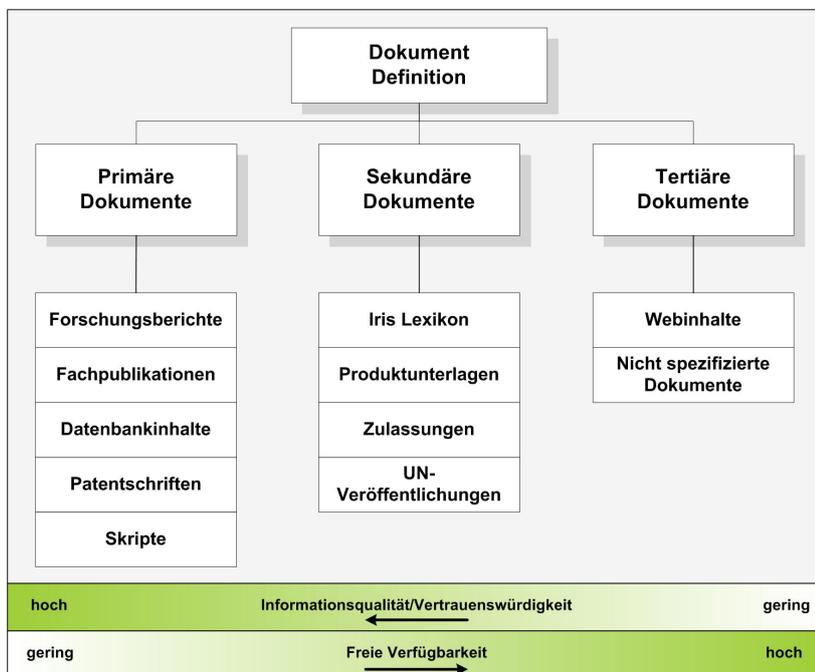


Abbildung 16: Definition der Dokumente und Zuordnung zu den definierten Qualitätsstufen

Dokumente sind die Basis für den Aufbau der holzbauspezifischen Systematik, der Querverbindungsmatrix sowie für das semantische Netz. Mit den genannten Dokumentquellen stehen unmittelbare und einfach zu beschaffende Dokumente mit entsprechender Qualität zur Verfügung. Der Anspruch an wissenschaftlich fundierte Daten und Informationen ist so erfüllt.

8.11 Zum Projektabschluss realisierte Kooperationen

Mit folgenden möglichen Partnern wurden Kooperationsgespräche durchgeführt:

- DIN/Beuth-Verlag, Berlin
- Deutsches Patent- und Markenamt (DPMA), und europäisches Patentamt (EPO), beide München
- Informationszentrum Raum und Bau (IRB), Stuttgart
- Technische Informationsbibliothek (TIB), Hannover
- Holzabsatzfonds / Informationsdienst-Holz / infoholz.de (IDH) Bonn
- Deutsche Gesellschaft für Holzforschung (DGfH), München
- Institut für Fenstertechnik (ift), Rosenheim

Der Austausch von hier relevanten Daten ist aus technischer Sicht möglich. Von den Partnern wird folgender Datenaustausch bevorzugt. Die Partner selektieren aus ihren Beständen die relevanten Datensätze und stellen dieses Paket IRIS zur weiteren Nutzung zur Verfügung. Die Bereitstellung wird in gewissen Zeitabständen aktualisiert. Treffer führen dann zurück auf den Partner. Die Partner sind nicht bereit, dass IRIS in den jeweiligen Systemen frei indexiert oder sucht. Unter dieser Voraussetzung ist die Bereitschaft zur Kooperation teilweise vorhanden.

Im Einzelnen:

Der *Beuth-Verlag* vertreibt Normen verstärkt über das Online-Portal.

Der Verlag ist aber nur bereit die holzbaurelevanten Datensätze zur Verfügung zu stellen, wenn dies von Seiten IRIS in Höhe von 3000,- € p.A. bezahlt wird. Lediglich eine Verlinkung ist kostenfrei möglich.

Das *DPMA/EPO* hat über das Portal DEPATISNET insgesamt 35 Mio. Patente und zugehörige Schriften online zugänglich gemacht. Das DPMA und das EPO haben aus ihren Erfahrung mit vergleichbaren Ansätzen zu IRIS geraten, lediglich auf die Möglichkeit eigenständiger Recherchen hinzuweisen.

Eine Indexierung / Suche über IRIS in den Datenbeständen wird weder gewünscht noch als sinnvoll eingestuft. Eine kostenfreie Verlinkung ist möglich.

Das *IRB* hat einer Kooperation zugestimmt und Datensätze zu holzbaurelevanten Forschungsberichten zur Verfügung gestellt. Zusätzlich haben sie alle bauaufsichtlichen Zulassungen in Kurzfassung freigegeben (die allgemeinen bauaufsichtlichen Prüfzeugnisse folgen).

Das *TIB* ist die zentrale deutsche Fachbibliothek Technik und hat einer Kooperation zugestimmt und wird Datensätze zu holzbaurelevanten Veröffentlichungen zur Verfügung stellen. Hierunter fallen auch die Datensätze des VIFAHOLZ und ELFIS, die zur Zeit umfangreichsten Datenbestände aus der Holz- und Forstwirtschaft. VIFAHOLZ und ELFIS werden komplett in TIB integriert.

Der *IDH* ist in Teilen integriert. Eine offizielle Freigabe steht noch aus und ist von einer noch zu treffenden Vereinbarung zur Erarbeitung und Pflege des Lexikon in IRIS abhängig.

Die *DGfH-Projektdatenbank* ist über das Portal infoholz.de zugänglich.

Hier sind noch technische Details zu klären.

Das ift hat noch sich noch nicht erklärt.

8.12 Qualitätssicherung

Wichtigster Baustein des Qualitätsmanagements und der Qualitätssicherung wird der Bereich der redaktionellen Systembetreuung sein. Der IRIS-Redakteur mit seiner Fachkenntnis garantiert die korrekte Auswahl und Zuordnung von Quellen und Dokumenten. Entsprechende Vorgaben und geeignete Instrumentarien sind erforderlich. Auf der anderen Seite wird die Qualitätssicherung durch die Einbindung der Kooperationspartner unterstützt. Der Benutzer kann durch gesetzte Attribute wie Hochschule, Institut oder allgemeine Webinhalte den für sich erforderlichen Qualitätsanspruch definieren.

8.13 Öffentlichkeitsarbeit

Für die Konzeption und laufende Projektumsetzung besteht ständiger Kontakt zu „externen Beratern“¹⁷. Hier sind Gespräche und die Teilnahme an Veranstaltungen, Seminaren und Tagungen zu nennen. Es wurden unterschiedliche, aber in das Portfolio des IRIS-Systems passende Experten angeschrieben. Alle durchgeführten und geplanten Aktivitäten sind auch als Marketing für IRIS und für den Technologietransfer zu sehen.

Auf der BDF Techniker-Tagung 2006 in Pilsen wurde das IRIS-Projekt erstmals einem breiteren Publikum vorgestellt. Mit Vertretern der Fertighausbranche war ein repräsentatives Spektrum an potentiellen Benutzern von IRIS vor Ort. Die Resonanz war erwartungsgemäß etwas verhalten, da es schwierig ist, ein System vorzustellen, welches man der Zielgruppe noch nicht „live“ präsentieren konnte, denn Vorführungen sind ein wesentlicher Beitrag jeder Überzeugungsarbeit.

TP 08 „Marketing“ führte auf der Tagung eine Umfrage durch. Auf die Frage: „Wird Ihnen eine internetbasierte Informationsplattform, welche als Dreh- und Angelpunkt in der Holzbaubranche Inselinformationen zentral verknüpft, die Recherchearbeiten verkürzen?“ antworteten 96% der Befragten mit „Ja“. Diese Umfrage und die bisherigen Erfahrungen zeigen, dass das ausgearbeitete Konzept auf breites Interesse stößt. Daneben halten es alle für sinnvoll, Kooperationspartner zur inhaltlichen Unterstützung einzubinden.

Die bisherigen Aktivitäten haben gezeigt, dass von Seiten der Unternehmen, Institute, Verbände und Hochschulen großes Interesse an der Informationsplattform besteht. Jedoch halten sich alle etwas zurück, wenn es um Kooperationspartnerschaften geht. Hier möchte keiner in die Rolle des Vorreiters treten. Wesentlicher Grund ist, dass gegenüber neuen Technologien und der Bereitstellung von Informationen noch eine erhebliche Skepsis besteht. Diese gilt es im weiteren Verlauf des Projekts abzubauen.

¹⁷ Unter „externe Berater“ fallen Unternehmen, Hochschulen, Verbände und Projektpartner.

9 Fachsystematik – Holzbauspezifische Systematik

Für das Informationssystem IRIS ist zur fachlichen Erschließung und Darstellung der Thematik „Holz- und Holzhausbau“ eine Fachsystematik erforderlich. Ziel der entwickelten Fachsystematik „IRIS“ ist es, die Fachterminologie, die Fachgebiete und den Fachjargon für die unterschiedlichen Benutzergruppen ganzheitlich zu erschließen.

9.1 Begriffsdefinitionen

In vielen Bereichen der Wissenschaft und Gesellschaft begegnet man dem Ausdruck „System“, der in den meisten Fällen undifferenziert für eine Vielzahl von Zusammenhängen verwendet wird. Auch die damit zusammenhängenden Begriffe „Systematik“, „Struktur“, „Modell“ und „Klassifikation“ werden oft - ohne auf die genaue Definition einzugehen - verwendet. Im Folgenden werden die genannten Begriffe definiert. Dazu wird sowohl auf Sichtweisen der Systemtheorie als auch auf klassische Wortdefinitionen, wie sie in Enzyklopädien zu finden sind, zurückgegriffen.

Tabelle 4: Zusammenstellung relevanter Begriffsdefinitionen zur Fachsystematik

Begriff	Definition
Struktur	Gegliederte Zusammensetzung aus <i>Elementen</i> und die Menge der <i>Relationen</i> , welche die Elemente miteinander <i>verknüpfen</i> , die sich innerhalb eines Systems herausbilden und diesem eine gewisse Stabilität ermöglichen.
System	Die <i>Gesamtheit</i> von materiellen oder geistigen Objekten, deren wesentliche Elemente so aufeinander bezogen sind, dass sie eine <i>Einheit</i> ergeben. Systeme organisieren und erhalten sich durch Strukturen.
Systemgrenze	<i>Abgrenzung</i> zu anderen bestehenden Systemen. Systeme werden von ihrer <i>Umgebung</i> bzw. von <i>Subsystemen</i> abgegrenzt.
Systematik und Klassifikation	<i>Planmäßige Darstellung</i> von Elementen, Klassen, Kategorien oder anderen abstrakten Konzepten eines Systems.
Modell	<i>Vereinfachtes Abbild</i> oder Repräsentation eines realen Systems.
Kategorie	<i>Menge von Objekten</i> , die meist <i>aufgrund gemeinsamer Merkmale</i> in Abgrenzung zu anderen Objekten zu einer Gruppe oder Klasse <i>zusammengefasst</i> werden.
Komplexität	Komplexität spiegelt sich in der <i>Anzahl</i> gleicher bzw. verschiedener <i>Elemente</i> , der Anzahl der gleichen bzw. verschiedenen <i>Beziehungen</i> sowie in der Anzahl der <i>Ebenen</i> wieder. Komplexität kann im Laufe der <i>Evolution</i> zunehmen.
Kompliziertheit	Schwierig nachvollziehbar, verstehbar, verständlich usw. Kompliziertheit ist <i>immer subjektiv</i> , sie ist deshalb abhängig von Vorkenntnissen, Vorurteilen, aber auch von Emotionen etc. Sie ist über den Verstehensprozess bzw. Lernprozess zeitlich veränderlich. Dies bedeutet, dass mit zunehmendem Verstehen eines Systems die Kompliziertheit abnimmt.

9.1.1 Möglichkeiten der Strukturierung

Für die Systematisierung und Strukturierung eines Themenkomplexes stehen mehrere Varianten zur Verfügung. Ziel ist es, die Komplexität zu reduzieren und die System-/Modellgrenzen zu definieren. Die Unterschiede liegen in der Möglichkeit und Art der Beziehungsdarstellung. Die wichtigsten sind:

(Mono-) Hierarchie

Bei einer Monohierarchie besitzt jedes Element - mit Ausnahme des obersten Elements - nur ein übergeordnetes Element, so dass die gesamte Klassifikation eine Baumstruktur besitzt.

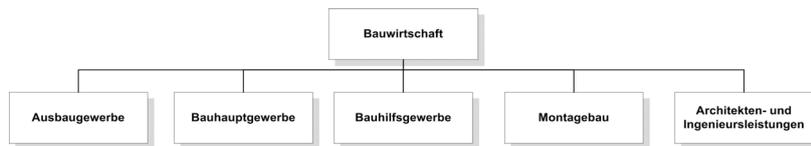


Abbildung 17: Hierarchie

Polyhierarchie (Multihierarchie)

In einer Polyhierarchie kann jede Bezeichnung mehrmals erscheinen, weil jeder Bezeichnung ein Begriff zugeordnet ist, der unter verschiedenen Aspekten an mehreren Systemstellen auftauchen kann.¹⁸

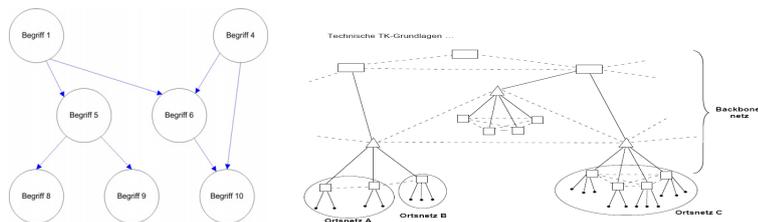


Abbildung 18: Polyhierarchie, bspw. Darstellung eines Telefonnetzes (rechtes Bild)

Vernetzung

Die Vernetzung erlaubt es jedem Element mit jedem anderen in Kontakt zu treten.

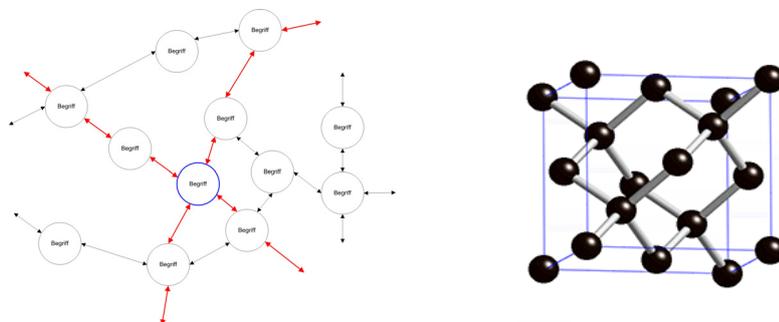


Abbildung 19: Vernetzung, bspw. Diamantstruktur (rechtes Bild)

¹⁸ Institut für Bibliotheks- und Informationswissenschaft, Humboldt Universität Berlin, www.ib.hu-berlin.de, Abruf vom 29.06.06

9.1.2 Vorhandene Systematiken

Zunächst ist zu analysieren, welche Systematiken verwendet werden. In Tabelle 5 sind einige wichtige Datenbanken und Systematiken zusammengefasst. Neben dem Holzbereich werden auch einige andere Fachgebiete berücksichtigt, um deren Erschließungskonzept eines Themas zu untersuchen.

„Eine Systematik soll nicht die Wirklichkeit in ihrer Gesamtheit abbilden.“¹⁹ Die Systematik muss jedoch empirisch begründet sein.²⁰ Dies kann für die Entwicklung einer neuen Systematik realisiert werden, indem bereits eingeführte und bewährte Systematiken als Anhaltspunkte dienen.

Auf Grundlage der obigen Ausführungen und Analysen wird nun eine Fachsystematik für den Holzbau formuliert. Für die Entwicklung der holzbauspezifischen Systematik wurden die folgenden Systematiken analysiert und auf die Übertragung und Anwendbarkeit hin geprüft:

- Systematik IHD (Holzbank)
- Systematik Informationsdienst Holz
- Systematik Holzabsatzfonds
- Systematik Dataholz

Jedes dieser Systeme hat einen interessengeleiteten Autor und ist damit subjektiv. In diesem Sinne ist auch die Fachsystematik in IRIS subjektiv. Für die Fachsystematik in IRIS erfolgte eine Orientierung an bestehenden Systematiken.

Tabelle 5: Zusammenstellung vorhandener Systematiken aus unterschiedlichen Bereichen

Art	Name	Inhalt	Zweck der Klassifikation	Art der Klassifikation
Online-Enzyklopädie	Wikipedia	Enzyklopädie mit über 400.000 Artikeln zu jedem Fachgebiet. Holzbau nur am Rande	Zuordnung von Dokumenten zu bestimmen Kategorien	Multihierarchie / Netzwerk
Klassifikation der Wirtschaftszweige	Destatis	Erfassung Branchenstruktur der Deutschen Wirtschaft. Einordnung Holzbau unter Baugewerbe	Systematisierung der Wirtschaftszweige zum Zweck statistischer Erhebungen	Hierarchie
Baublauf		Arbeitsfelder des Planers vor/während der Bauausführung, Zielgruppe Architekt/Planer. Nicht speziell für Holzbau, aber in weiten Teilen übertragbar.	Chronologische Anordnung planerischer Tätigkeiten im Bauverlauf	Zeitskala
Bibliothek	Regensburger Verbundklassifikation	Klassifikationssystem an wissenschaftlichen Bibliotheken. Zielgruppe Bibliotheksbenutzer. Einordnung Holzbau unter Architektur, Bauingenieurwesen, Handwerk und Gewerbe	Zuordnung von Dokumenten zu bestimmen Kategorien	Hierarchie
Datenbank	Dataholz	Bauphysikalisch, ökologisch geprüfte und/oder zugelassene Holz- und Holzwerkstoffe, Baustoffe, Bauteile, Bauteillanschlüsse. Zielgruppe Planer für Nachweisführung gegenüber Baubehörden. Speziell für Holzbau	Zuordnung von Bauteilen zu bestimmten Kategorien	Hierarchie
Datenbank	IHD	Fachlich aufbereitete Literaturnachweise (1986 bis heute). Zielgruppen: Holzforschung, Holz-, Möbel- und Zulieferindustrie. Holzbau nur am Rande	Zuordnung von Dokumenten zu bestimmen Kategorien	"Vernetzte Hierarchie"
Datenbank	Informationsdienst Holz	Holzbau Handbuch; Gliederung dreistufig (Reihen, Teile, Folgen), ständig erweitert. Inhalte werden mit Fachkreisen abgestimmt und regelmäßig überarbeitet. Zielgruppen: Architekten, Tragwerksplaner, Betriebe, Baubehörden, Dozenten und andere Baufachleute Abdeckung großer Bereiche des Holzbaus.	Zuordnung von Dokumenten zu bestimmen Kategorien	Hierarchie
Klassifikation der Holzwirtschaft	HAF	Systematisierung nach Branchenstruktur der Deutschen Holzwirtschaft. Holzbau wird Zimmerer und Holzbaugewerbe sowie unter Fertigbau zugeordnet.	Systematisierung der Holzwirtschaft zum Zweck statistischer Erhebungen	Multihierarchie
Klassifikation der Warenwirtschaft	UNSPSC	United Nations Standard Products and Services Code: International eingesetztes Klassifikationssystem der Warenwirtschaft. Unternehmenübergreifende Klassifikation von Waren und Dienstleistungen aller Art.	Schnelle, einfache u. akkurate Klassifikation von Gütern	Hierarchie
WHO Krankheitenklassifikation	ICD	Systematisierung nach Krankheiten und verwandten Gesundheitsproblemen. Zielgruppe Ärzte und Krankenhäuser	Erfassung aller Krankheiten in einer einheitlichen Systematik	Hierarchie

Durch die Verwendung etablierter Fachsystematiken wird die holzbauspezifische Systematik auf einen breiten Konsens gestellt. Diese etablierten Systematiken werden zum Zwecke einer neuen Systematisierung erweitert, angepasst und zusammengefasst. Es wird eine Fachsystematik geschaffen, die dem *Vorverständnis möglichst vieler Fachleute und Beteiligte* im Holzbau gerecht wird.²¹ Neben der Objektivierung der holzbauspezifischen Systematik

¹⁹ Vgl. Jürgen Plieninger: Umsystematisieren; Berufsverband Information Bibliothek / Kommission für One-Person Libraries (Hrsg.), 1. Aufl., 2005

²⁰ G. Ropohl, Allgemeine Technologie – Eine Systemtheorie der Technik, 2. Aufl., Leipzig 1999

²¹ Vgl. Jürgen Plieninger: Umsystematisieren; Berufsverband Information Bibliothek / Kommission für One-Person Libraries (Hrsg.), 1. Aufl., 2005

unterstreicht eine Zusammenführung der unterschiedlichen Systematiken den Charakter der unabhängigen, vernetzenden Informationsplattform IRIS. Die *Verfeinerung* der Systematik erfolgte im weiteren Projektverlauf durch Einbindung der betroffenen Personenkreise. Hier ist die Zusammenarbeit mit geeigneten Projektpartnern zu nennen.

9.2 Fachsystematik Holzbau

Der Holzbau steht mit vielen Themengebiete innerhalb des Bauwesens in mittelbarer oder unmittelbarer Beziehung. Die Systemgrenze – und damit die *inhaltliche Priorität* und *Fokussierung* – wird im Projekt um den Themenkomplex Holzbau gezogen. Dieser Bereich wird primär systematisiert. Eine Systematisierung und Einbindung aller angrenzenden Themenkomplexe im Bauwesen würde den Projektrahmen sprengen.

Das Themengebiet bzw. die Fachsystematik „Holzbau“ steht in der Holzbranche in vielseitigen Wechselbeziehungen. Die Holzbranche²² lässt sich unterteilen in folgende Hauptkategorien:

- Holzbau
- Fenster- und Fassadenbau
- Innenausbau und Trockenbau
- Konstruktiver Ingenieurbau
- Forstwirtschaft
- Zellstoff- und Verpackungsindustrie
- Sägeindustrie
- Holzverarbeitung und Möbelindustrie
- Sonstige Holzanwendungen

In der Analysephase wurden die angrenzenden Themengebiete bzgl. deren Inhalte untersucht. Es konnten hierbei gleiche/ähnliche, überlappende Kategorien festgestellt werden. Auch mit anderen Bereichen des Bauwesens wie Mauerwerksbau, allg. Ingenieurbau und Stahlbau gibt es Berührungspunkte. Die Gemeinsamkeiten bzw. Schnittmengen wurden in Unterkategorien zusammengefasst. Diese spiegeln sich in allen Themenbereichen der Holzbranche, aber auch des Bauwesens, wieder. Eine vorläufige Fachsystematik ist in Abbildung 21 dargestellt.

Aus Gründen der Übersichtlichkeit und dem aktuellen Kenntnisstand folgend wurde für die Hauptkategorie „Holzbau“ auf eine Differenzierung nach Zimmerer, Holzbaugewerbe und Fertigbau verzichtet. Zur Vereinfachung der Darstellung wird die Systematik zunächst grob unterteilt nach:

Hauptkategorie – Unterkategorie – Themengebiet

Alle Themengebiete sowie Haupt- und Unterkategorien sind untereinander vernetzt. Ausgangsbasis für die Hauptkategorien ist die Fachsystematik des Informationsdienstes Holz. Für die Systematisierung der Unterkategorien wurde sowohl die Fachsystematik des IHD als auch die Systematik von Dataholz aufgegriffen. Zur Vervollständigung der Unterkategorien erfolgten entsprechende Erweiterungen:

- Technologie (um der fortschreitenden Automatisierung des Holzbaus durch moderne Maschinentechologie Rechnung zu tragen)
- Betriebswirtschaft
- Umweltschutz – Abfallwirtschaft – Ökologie

²² Holzabsatzfonds, Die deutsche Holzwirtschaft, Zahlen und Fakten, Bonn 2005

- Normen und Gesetze
- Gebäudetechnik

Die hier entwickelte und in Abbildung 21 skizzierte Systematik stellt eine allgemeine Verständigungsbasis für die Experten der verschiedenen Spezialgebiete, die bei komplexen Projekten zusammenarbeiten, dar und klärt diese über die Zusammenhänge auf, in denen ihr Spezialgebiet steht. Im übertragenen Sinne dient die Systematik dazu, das komplexe Gebilde des Bauwesens und Holzbaus vereinfacht darzustellen. Das heißt, dass das reale System Holzbau so vereinfacht und strukturiert in einem Modell abgebildet werden muss, dass alle wesentlichen Themenbereiche behandelt werden und gleichzeitig ein Überblick über die im Holzbau relevanten Wissens- und Themenbereiche gegeben wird. Das reale System „Holzbau“ ist und bleibt jedoch komplex.

Mit der Umsetzung der Fachsystematik Holzbau in IRIS werden folgende Ziele verfolgt:

- *Strukturierte* Abbildung des realen Systems Holzbau in einem Modell, so dass alle wesentlichen Themenbereiche *lückenlos erfasst* werden und gleichzeitig ein Überblick über die im Holzbau relevanten Wissens- und Themenbereiche gegeben wird.²³

Des Weiteren:

- Problemlose *Erweiterbarkeit* der Systematik im Laufe des Projektes, etwa um neue Themengebiete nachträglich aufnehmen zu können
- Vereinfachung der Erschließung und Bearbeitung des Themenkomplexes Holzbau, sowohl für den Anwender als auch für den IRIS-Redakteur
- Unterstützung der Navigation durch übergeordnete Strukturen (Kategorien, Themengebiete)
- Unterstützung des „lernenden“ Programmverhaltens
- Zuordnung von Dokumenten zu Kategorien oder Themengebieten
- Verbesserung der Auffindbarkeit von Dokumenten in Themengebieten
- Statistische Auswertungen und Grobanalysen wie z. B. „Wie viele Dokumente sind zum Themengebiet Schallschutz verfügbar?“

In erster Näherung lässt sich die Systematik hierarchisch darstellen. Bei genauer Betrachtung benötigt man für die ganzheitliche Erschließung einen mehrdimensionalen Raum (Abbildung 20, hier 3-dimensional). Während bei der Hierarchie die Gliederungsstufen vorgegeben sind, bleibt der Pfad im n-dimensionalen Raum offen.

²³ Mit dieser Entwicklung verbunden ist die Zielsetzung, eine höhere Striktheit und Normiertheit der Abläufe zu erreichen, um so eine bessere planerische Durchdringung aller relevanten Prozesse sicherzustellen. Mit dieser Entwicklung einhergehend ist eine verbesserte Transparenz. Vgl. F. J. Radermacher: Komplexe Systeme und lernende Unternehmen, 20. AUT-Kolloquium Wissensmanagement der Siemens AG, Nürnberg 1996, S.427

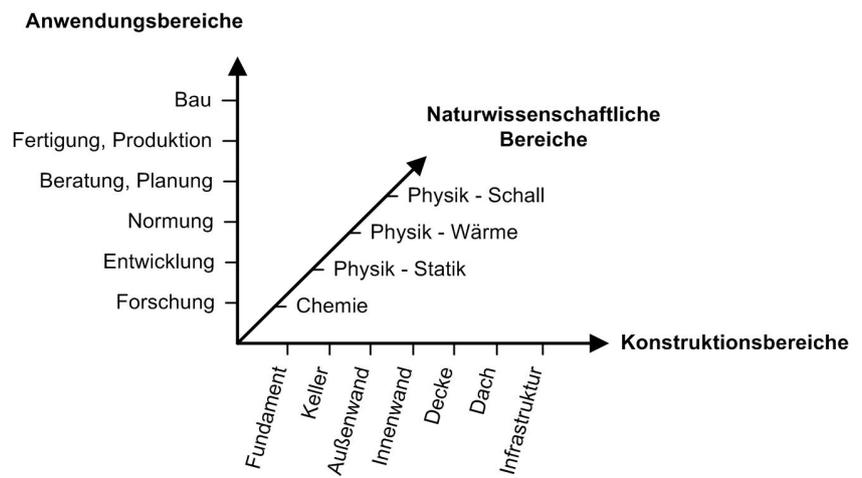


Abbildung 20: Die Fachsystematik Holzbau im mehrdimensionalen Raum

Mit der entwickelten holzbauspezifischen Systematik können alle Bereiche des Holzbaus - und der Themenkomplexe, die mit dem Holzbau in Verbindung stehen - abgedeckt werden.

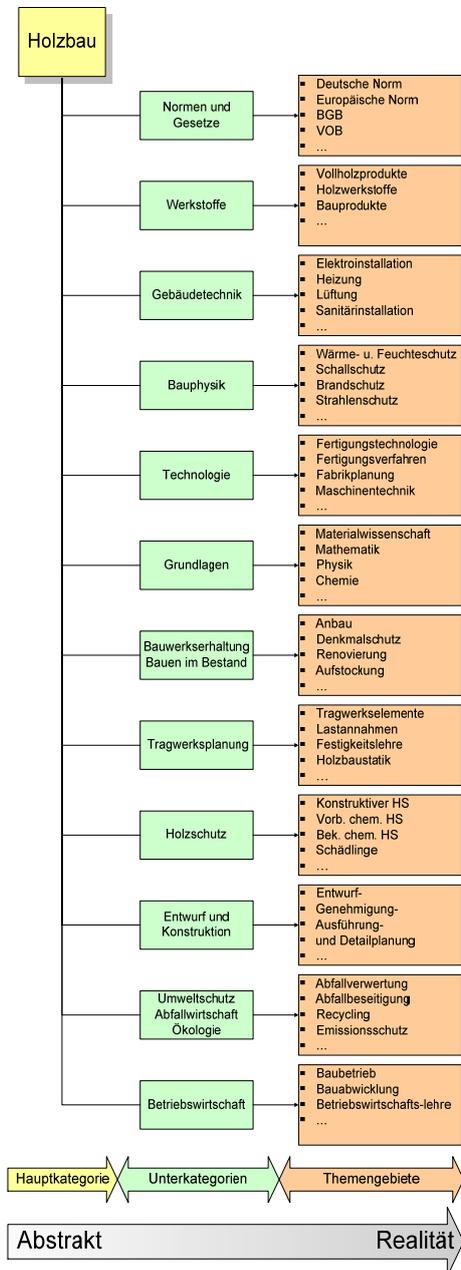


Abbildung 21: Schema holzbauspezifische Systematik

9.3 Semantisches Netz

Mit der zuvor erarbeiteten Systematik ist die Strukturierung des Holzbaus auf die Grundelemente gegeben. Es wurde festgestellt, dass es nicht möglich ist, mit dieser Methode/Art der Strukturierung alle Beziehungen innerhalb des Holzbaus hinreichend darzustellen. Ferner sind die Beziehungen zu anderen Themen des Bauwesens schwer aufzuzeigen und führen schnell zu Unübersichtlichkeit. An dieser Stelle bietet sich der Einsatz von semantischen Netzen an.

Die holzbauspezifische Systematik wird innerhalb des IRIS-Systems in einem *semantischen Netz*²⁴ abgebildet. Wesentliches Merkmal semantischer Netze ist, dass neben *Begriffen* auch benannte gerichtete *Relationen* (im Sinne von „Teil von“, „umfasst“ etc.) erfasst werden. Dadurch können komplexe Systeme realitätsgetreuer abgebildet werden als mit traditionellen Methoden. In IRIS wird die holzbauspezifische Systematik als *Ordnungskriterium* zugrunde gelegt, um eine klare Klassifizierung und Indexierung von Informationen für den Holzbau zu ermöglichen.

Da es nicht möglich ist, das semantische Netz als Ganzes graphisch abzubilden, wird beispielhaft ein Ausschnitt des semantischen Netzes dargestellt. Ein konkretes Beispiel hierzu aus dem Bereich Bauphysik, die Suchanfrage an das System könnte so aussehen: „Schallschutz Wand Dach“. Geliefert wird im semantischen Netz das in Abbildung 22 skizzierte Ergebnis.

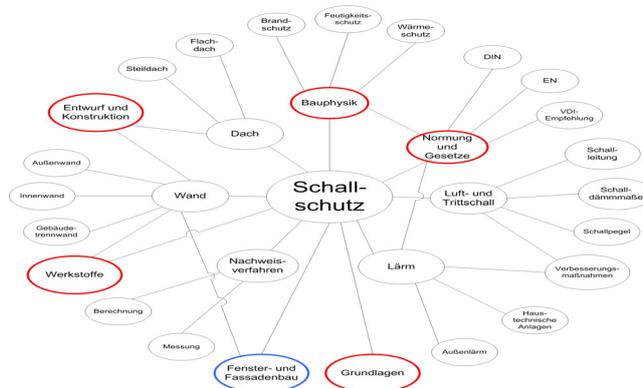


Abbildung 22: Schematischer Ausschnitt aus dem semantischen Netz

Das Element (Themengebiet) „Fenster- und Fassadenbau“ hat in der Systematik zwar keine direkte Verbindung mit der Suchanfrage „Schallschutz Wand Dach“, jedoch wird die mittelbare Verbindung über Relationen der Art „Fenster ist Teil von Wand“ oder „Schallschutz steht in Beziehung zu Fenster“ hergestellt.

²⁴ Siehe Kapitel 5.5 Semantisches Netz

Im semantischen Netz werden Begriffe als Knoten dargestellt und Beziehungen als gerichtete Kanten, die jeweils paarweise Knoten miteinander verbinden. Als Knoten bezeichnet man in der Graphen-Theorie ein Element der Knotenmenge eines Graphen. Graphen enthalten neben der Knotenmenge eine Kantenmenge, die beschreibt, welche Knoten paarweise mit welcher Beziehung in welche Richtung verbunden sind.^{25 26} Ferner gibt es eine Ergebnisliste, in der die relevanten Dokumente für die Suchbegriffe nach ihrer Relevanz geordnet angezeigt werden.

Die Menge der Begriffe innerhalb der Systematik bzw. des semantischen Netzes bildet ein kontrolliertes Vokabular, die sog. *Fachterminologie*²⁷. In dieser Fachterminologie ist es möglich, über bestimmte Relationen Wortgruppen (Syntagma), Wortformen (Kollokationen), Synonyme, Antonyme etc. festzulegen. Die Beschreibung der Gesetzmäßigkeiten von Fachterminologien und die Erarbeitung der Fachterminologie von Fachsprachen sind Gegenstand der Terminologie-Lehre. Für die Entwicklung oder Entdeckung der Fachbegriffe selbst werden bisher kaum automatisierte Verfahren eingesetzt. Als Grundlage für die Erstellung der Fachterminologie dienen Fachtexte (Forschungsberichte etc.) sowie ein allgemeinsprachlicher Textkorpus, aus denen die Terme mittels Indexierung extrahiert werden.

Die Verfeinerung der Systematik ist unabhängig vom Themengebiet möglich, da lediglich ein neuer Knoten (zur Erweiterung eines Themengebietes) im semantischen Netz eingefügt und mit dem bestehenden Netz verknüpft werden muss. So wird z. B. für ein Dokument, das sich mit Glasverklebungen im Automobilbau beschäftigt, ein Knoten „Automobilbau“ angelegt und dieser dann anschließend mit den Bereichen „Fenster- und Fassadenbau“, „Verklebung“, „Verbindungstechnik“, „Werkstoffe“ etc. entsprechend verknüpft.

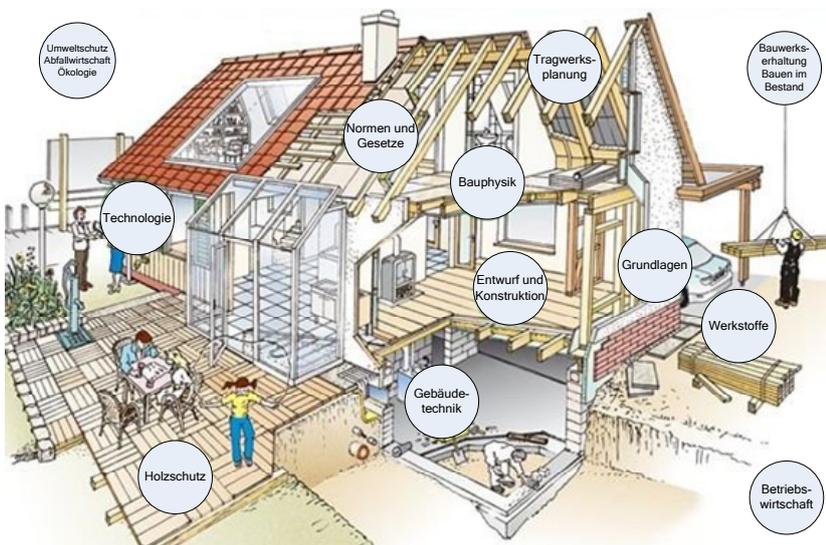


Abbildung 23: Holzbauspezifische Systematik am Beispiel Holzhaus. (Quelle: Hagebau Holzhandel)

²⁵ www.wikipedia.de, Knoten (Graphentheorie), Abruf vom 29.06.06

²⁶ www.wikipedia.de, Kante (Graphentheorie), Abruf vom 29.06.06

²⁷ Hier finden sich wesentliche Aspekte der Strukturierung des Fachwissens (Sprachnormierung).

10 Informationstechnische Konzeption und Realisierung

10.1 Grundlegende Betrachtungen

Ein „Integriertes relationales Informationssystem für den Holzbau – IRIS“ stellt ein komplexes Produkt dar, das im Wesentlichen folgendermaßen strukturiert ist:

- Es gibt mehrere spezialisierte Systemkomponenten, die unter anderem hinsichtlich Benutzeroberfläche und Bedienung vereinheitlicht sind.
- Diese harmonisierten Komponenten sind in einem Gesamtsystem zusammengeführt.
- Aufgabenschwerpunkte sind Informationsgewinnung, -sammlung, -konzentration und -abfrage.
- Beziehungen (Relationen) zwischen Informationen werden abgebildet und erweitern die Suche mit entsprechenden Navigationsmöglichkeiten.
- Das Gesamtsystem ist auf das Fachgebiet Holzbau ausgerichtet.

Es gibt bzw. ergibt sich eine Reihe von Anforderungen an ein derartiges System. Nachfolgend werden die technischen Bestandteile und deren Realisierungen dargestellt und erläutert.

10.1.1 Aus Sicht der fachlichen Anwendung

- Ein wesentlicher, zentraler Punkt ist die Abbildung einer holzbauspezifischen Systematik.
- Das System und die darüber zur Verfügung gestellten Informationen sind allgemein zugänglich.
- Es sind unterschiedliche Benutzer und eine entsprechende Verrechnung zu berücksichtigen.
- Alle Informationen sind einfach und kurzfristig aktualisierbar.
- Die Oberfläche soll zweisprachig gestaltet sein. Die primäre Sprache ist Deutsch, die sekundäre Sprache Englisch.
- Das System ist mehrbenutzerfähig, robust und leistungsstark.
- Das System soll derart realisiert werden, dass so wenig Metadaten wie möglich verwendet bzw. verwaltet werden müssen. Ziel ist es, den Verwaltungsaufwand und mögliche Fehlerquellen zu reduzieren und das erforderliche Expertenwissen für die Eingabe gering zu halten.

10.1.2 Aus Sicht der Informationstechnik

- Die erforderliche technische Ausstattung auf der Seite der Anwender entspricht dem aktuellen Stand der Informationstechnik und der Ausstattung, die einem Anwender normalerweise zur Verfügung stehen. Konkret wird vom Zugang zum Internet und der Verwendung eines üblichen Webbrowsers ausgegangen.
- Das Informationssystem soll – soweit es sinnvoll und realisierbar ist – portabel sein. Davon betroffen sind insbesondere Betriebssystem, Programmierung, Daten, Server und Client. Es ist wartbar sowie erweiterbar und mit Mitteln sowie Methoden realisiert, die dem aktuellen Stand der Informationstechnik entsprechen.
- Bewährtes soll zur Realisierung der geforderten Funktionalität genutzt werden, wenn dadurch die Entwicklung vereinfacht, beschleunigt und technologisch aktuell gehalten werden kann. Insbesondere ist das die Verwendung von geeigneten verfügbaren, gepflegten und zukunftsfähigen Komponenten und Modulen, deren Schnittstellen und Quelltexte offen gelegt sind, die durch Veränderung des Quelltextes oder der Konfiguration angepasst werden können und bei deren Verwendung keine Lizenzkosten entstehen.
- Da der Zugang zum Informationssystem über das Internet und damit über einen ungesicherten Kanal erfolgt, soll die Kommunikation selbst gesichert erfolgen.

10.2 Architektur und Plattform

Aufgrund der Anforderungen bietet sich zunächst die Verwendung einer Client-Server-Architektur an. Da auf der heterogenen Anwenderseite nur minimale Voraussetzungen gemacht werden können, kommt hier – entsprechend dem Stand der Technik – nur ein Webbrowser ohne Verwendung von Plug-Ins in Frage.

Mit der Basis Internet steht die Protokoll-Familie TCP/IP fest. Webbrowser bieten eine Oberfläche, die Hypertext unterstützt. Die Kommunikation zwischen Client und Server erfolgt über HTTP. Die Inhalte werden entsprechend der Seitenbeschreibungssprache HTML formuliert. Von der Clientseite wird somit lediglich die Bereitstellung der Ein- und Ausgabefunktionalität erwartet. Webbrowser bieten hier reichhaltig Funktionalität und Komfort.

Maschine-Maschine-Schnittstellen werden insbesondere für die Indexierung benötigt. Weitere werden gegebenenfalls für Web-Services u. ä. nötig sein.

10.2.1 Betriebssysteme

Bei ausschließlicher Nutzung des Informationssystems über Webbrowser spielt das client-seitige Betriebssystem keine Rolle, sofern für das jeweilige Betriebssystem aktuelle Browser verfügbar sind. Häufig verwendete Betriebssysteme sind beispielsweise MS Windows, Mac OS und Unix-artige Systeme.

Die Serverseite muss dynamisch generierte Inhalte liefern und über HTTP bereitstellen können.

Erforderlich sind Mittel zur Datenhaltung sowie Software, die diese Daten auf client-seitige Anforderung aufbereiten und ausliefern kann. Die Datenhaltung kann über eine Datenbank bzw. speziell organisierte Strukturen im Dateisystem erfolgen. Geeignete Software, die durch Konfiguration angepasst werden kann, existiert gegebenenfalls oder muss durch individuelle Programmierung erstellt werden.

10.2.2 Programmierung

Entsprechend den Vorgaben ist die Verwendung bzw. Entwicklung eines Webservers erforderlich. Hierfür stehen insbesondere folgende Alternativen zur Verfügung:

10.2.2.1 CGI (Common Gateway Interface)

Das CGI wird von einer speziellen Webserver-Software (z. B. Apache HTTP Server oder MS IIS) umgesetzt, die im Wesentlichen zwei Interfaces zur Verfügung stellt. Einerseits übernimmt sie die Kommunikation mit dem Browser und andererseits die Kommunikation mit der Systemlogik.

Vorteile:

- Es ist die einfachste Möglichkeit, dynamische Webinhalte zu generieren.
- Man kommt schnell zu einem ersten Ergebnis.
- CGI ist für alle gängigen Betriebssysteme verfügbar.

Nachteile:

- Jede Seitenanfrage des Clients generiert einen neuen Prozess im Betriebssystem des Servers. Da dieser Vorgang zeitaufwändig ist, führt das insgesamt zu einer schlechten Performance.
- Abgesehen von sehr einfachen Anwendungen ist es erforderlich, dass mit Sitzungen gearbeitet wird, insbesondere wenn eine Benutzeranmeldung am System zu realisieren ist. Mit HTTP selbst kann das jedoch nicht erreicht werden, da es nicht vorgesehen ist, eine Verbindung über die Auslieferung einer Seite hinaus aufrecht zu erhalten. Das Merken von Zuständen (insbesondere Sitzungsdaten) ist mit CGI schwierig und aufwändig.
- Sicherheitsprobleme

10.2.2.2 Speziell für Webserver ausgerichtete Sprachen

In dieser Kategorie findet man insbesondere PHP²⁸.

Vorteile:

- verfügbar für alle gängigen Betriebssysteme
- spezialisiert auf Webserver-Funktionalität
- umfangreiche Funktionalität
- weit verbreiteter Einsatz
- leicht erlernbar

Nachteile:

- vielfach sind die vorhandenen Konzepte nicht durchgängig
- objektorientierter Ansatz, der jedoch nicht vollständig ist; vieles ist noch prozedural
- Skriptsprache

²⁸ Rekursives Akronym für: PHP Hypertext Preprocessor

10.2.2.3 Hochsprache

Skriptsprachen sind in der Regel für kleine und schnell zu erstellende Anwendungen konzipiert. Sie werden mit höheren Programmiersprachen erstellt und erfahren dabei Einschränkungen. Die direkte Nutzung höherer Programmiersprachen bietet somit Vorteile, z. B. hinsichtlich Performance, Flexibilität und Funktionalität. Die Erstellung von Software in Skriptsprachen ist bei kleinen Programmen effektiver, mit zunehmender Komplexität steigt jedoch der Aufwand im Vergleich zu Hochsprachen überproportional stark an. Aktuelle höhere Programmiersprachen sind insbesondere C, C++, Java und C#. Ein entscheidendes Kriterium ist die Verfügbarkeit einer API für die Internetprogrammierung.

10.2.2.4 .NET-Technologie

.NET (sprich: „dot net“) ist eine Implementierung des Common-Language-Infrastructure-Standards. Die CLI stellt eine Basis zur Ausführung von Programmen dar, die mit unterschiedlichen Programmiersprachen erstellt wurden. Als gemeinsame Klassenbibliothek steht die so genannte Framework Class Library (FCL) zur Verfügung. Die Programme werden über eine virtuelle Maschine (ähnlich wie bei Java) ausgeführt.

10.2.2.5 Java

Unter Berücksichtigung der zahlreichen Anforderungen fiel die Wahl letztendlich auf Java. Java ist eine noch verhältnismäßig junge objektorientierte, plattformunabhängige Sprache, die bewährte Konzepte, insbesondere von C und C++, übernimmt und problematische Konzepte im Wesentlichen vermeidet. Es existiert ein großes Spektrum an weitgehend konsistenten effektiven und effizienten APIs für allgemeine wie für spezielle Aufgaben.

Java bietet vielfältige Unterstützung für verschiedene Arten von Anwendungen. Im Zusammenhang mit Webservern sind hier insbesondere Applets, Servlets und JSPs (Java Server Pages) von Interesse. Insofern liefert die Spezifikation einer Standardarchitektur für die Ausführung von Anwendungskomponenten und Diensten, die so genannte Java EE (Java Plattform, Enterprise Edition), einen interessanten Aspekt. Sie stellt einen allgemein akzeptierten Rahmen zur Verfügung, um auf dessen Basis aus modularen Komponenten verteilte, mehrschichtige Anwendungen entwickeln zu können. Ein Webserver lässt sich gemäß einer MVC2-Architektur (für Webanwendung spezialisierte Variante der Model-View-Controller-Architektur) mit einem so genannten Application-Server sowie Servlets (Controller) und JSPs (View) realisieren. Ziel von MVC ist ein flexibles und übersichtliches Software-Design, das Veränderungen, Erweiterungen und die Wiederverwendung einzelner Komponenten erleichtert, indem die Software in Datenmodell (M), Präsentation (V) und Programmsteuerung (C) gegliedert wird.

Ein weiterer wesentlicher Punkt ist die Verfügbarkeit von Software, die für die Realisierung von Teilsystemen verwendet werden kann. Darüber hinaus ist es notwendig, dass sich die Komponenten möglichst nahtlos zusammenfügen lassen. Mit *Eclipse* ist eine sehr leistungsfähige Entwicklungsumgebung verfügbar.

10.2.3 Verwendete Plattform

- Server-seitiges Betriebssystem: Linux; mögliche Alternativen: Unix-artige Systeme, MS Windows
- Programmiersprache: Java (Java EE)

- Application-Server: Jakarta Tomcat; mögliche Alternativen: JBoss u. a.
- Datenbank: MySQL; mögliche Alternativen: Oracle u. a.

10.3 Systemkomponenten – Übersicht

Das Informationssystem gliedert sich in die Systemkomponenten

- *Kernsystem*
Es dient der Integration und Administration der weiteren Teilsysteme, insbesondere Benutzerverwaltung, Login, Suche und Navigation, Konfiguration der Indexier- und Suchfunktionalität, Konfiguration des semantischen Netzes
- *Lexikon*
Darstellung von formatierten Texten und Bildern; Verwendung von Referenzen; Einbindung unterschiedlicher Dokumente
- *Indexierung, Suche*
Grundlage für das Auffinden von Informationen
- *Semantisches Netz*
Qualifizierte, gerichtete und gewichtete Verbindung von Informationen, deren graphische Darstellung sowie die begriffliche Navigation
- *Datenhaltung*
Erfolgt überwiegend in Datenbank-Tabellen, Indexierung über das Dateisystem (aus Performance-Gründen)

10.4 Indexierung

Indexierung ist, ganz allgemein, die Grundlage für die Bereitstellung von Inhalten, für die effiziente Suche sowie eine weitergehende Auswertung in einem Knowledge-Management-System. Das Einlesen von Dokumenten und die Suche laufen dabei asynchron ab, d. h. sie finden getrennt voneinander zu verschiedenen Zeitpunkten statt. Zum einen müssen dadurch die Datenquellen nicht jederzeit direkt verfügbar sein, zum anderen bringt es Vorteile bei der Suchgeschwindigkeit, entlastet die Datenquellen durch weniger Zugriffe und ermöglicht eine umfangreiche, intelligente Auswertung. Die Indexierung seiner Inhalte ist somit für das Informationssystem IRIS unverzichtbar.

10.4.1 Aufgabenstellung

Die Aufgaben, die sich für die Bewältigung des Prozesses „Indexierung“ ergeben, sind im Einzelnen:

- Bereitstellung eines Moduls („Crawler“) für das Durchlaufen des Suchraums (systematisches Absuchen von Datenquellen – lokal oder auf Fremdsystemen – und Auslesen von Dokumenten)
- Extrahieren und Aufbereiten der Inhalte aus diesen Dokumenten (z. B. Überschriften, normaler Text und Metaangaben aus MS Word-Dateien)
- Das eigentliche Indexieren dieser Inhalte und Schreiben eines so genannten „Invertierten Index“
- Die Bereitstellung grundlegender Suchfunktionen auf dem bzw. den Index/en

Der gesamte Vorgang läuft asynchron ab. Die Gliederung des Vorgangs ist in Abbildung 24 in 3 Ablaufschritten dargestellt.



Abbildung 24: Prinzipieller Ablauf von Indexierung und Suche

10.4.2 Recherche

Zu „Indexierung“ finden sich umfangreiche Theorien, die direkt oder indirekt damit zusammenhängen, vor allem Computerlinguistik (u. a. Syntaxanalyse, Wortstammanalyse, semantische Analyse), Informationstheorie, Such- und Ranking-Algorithmen.

Nach der Erarbeitung der Grundlagen war der nächste Schritt die Auswahl eines geeigneten Indexierungstools. Die freie Bibliothek „Lucene“ der Apache Software Foundation[®] gilt als leistungsfähiger Standard im Java-Umfeld. Sie bietet Funktionen zum eigentlichen Indexieren von reinem Text, zum Schreiben und Lesen eines Indexes sowie zum Verarbeiten von Suchanfragen mit eigener, umfangreicher Syntax. Das Zuführen von Text, also das Crawling, sowie bei der Suche die Auswertung und Präsentation der Treffer muss man noch selbst bewerkstelligen. Für das Durchlaufen des Suchraums bieten sich drei grundlegende Varianten an:

- Eigenentwicklung eines Crawlers
- Verwendung und gegebenenfalls Anpassung eines frei verfügbaren Crawlers
- Einsatz eines kommerziellen Produktes

Im Hinblick auf Flexibilität bezüglich Änderungen und Erweiterungen wurde die zweite Variante „Verwendung eines frei verfügbaren Crawlers“ gewählt. In die engere Auswahl kamen die Open-Source Projekte „Nutch 0.7.1“, „Red Piranha 0.3“ und „Regain 1.1“. Die Entscheidung zugunsten „Regain“ ist damit zu begründen, dass es das ausgereifteste Produkt ist und zudem bereits für eine ganze Reihe von gängigen Dateiformaten Dokumentenverarbeiter, „Präparatoren“ genannt, mitbringt. Enthalten sind fertige Präparatoren für die Formate MS Office (Word, Excel, Powerpoint), Adobe PDF, HTML, XML, OpenOffice und andere. Aufgrund der Anforderungen für einen Einsatz in IRIS sind Anpassungen nötig. Ein wesentlicher Punkt, der für die Verwendung unter IRIS relevant ist, ist die Möglichkeit, auch das sich laufend ändernde IRIS-Lexikon zeitnah in einen Index aufzunehmen. Nach einigen Versuchen und Analysen hat sich ergeben, dass das für das Lexikon gewählte Produkt JSP-Wiki es auf relativ einfache Weise zulässt, seine Inhalte nicht im Dateisystem, sondern laufend aktualisiert in einer Datenbank abzulegen. „Regain“ kann diese mit Anpassungen auslesen und so mit der Software für die Enzyklopädie zusammenarbeiten.

Während der Laufzeit des Projektes wurde auch „Regain“ ständig weiterentwickelt. Eine Analyse hat ergeben, dass in der gegen Ende der Projektlaufzeit aktuellen Version 1.2.3 einige Fehler und Schwächen korrigiert worden sind und außerdem eine neuere Version von „Lucene“ (2.2.0 statt 2.0.0) zum Einsatz kommt. Ein Umstieg erschien daher sinnvoll, obwohl etliche Anpassungen bzw. Erweiterungen an „Regain“, die extra für IRIS vorgenommen worden waren, hierfür überarbeitet werden mussten. Die in IRIS eingesetzte Software ist somit „Regain“ in der Version 1.2.3.

10.4.3 Probleme

Beim Indexieren gibt es eine Reihe von spezifischen Schwierigkeiten, die insbesondere die Qualität von Suchergebnissen beeinträchtigen können. Dies sind z. B. nicht lesbare ältere Dateiformate wie MS Word 97, der Umgang mit Duplikaten, d. h. exakt gleichen Dokumenten aus verschiedenen Quellen, falsch oder unsinnig belegte Metadaten etwa in aus anderen Formaten generierten PDF-Dateien oder die Behandlung von Grafiken in Dokumenten.

10.4.4 Integration

Indexierung ist eine zentrale Querschnittsfunktion mit Verbindungen zum Crawler, zum Sucher-Modul, dem semantischen Netz (siehe Kapitel 10.5), dem Lexikon (siehe Kapitel 10.6) und dem Lernen (siehe Kapitel 10.7).

10.4.5 Das Crawling

Die Arbeitsweise des Crawlers kann man Abbildung 25 entnehmen. Hier dargestellt ist der prinzipielle Arbeitsablauf des Crawling.

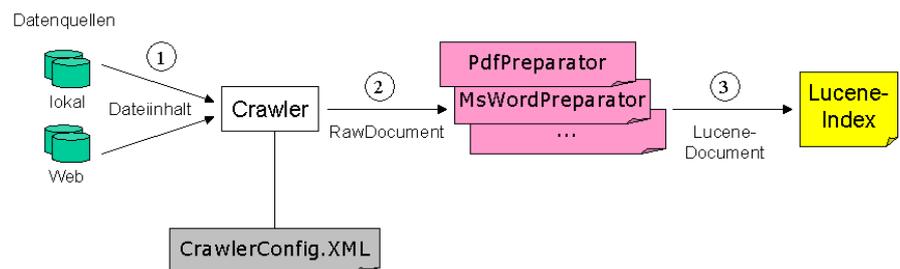


Abbildung 25: Crawling-Prozess in Regain

Das Crawler-Modul greift anhand seiner Konfiguration (in obiger Abbildung: „CrawlerConfig.xml“) auf die gewünschten Datenquellen zu: das lokale Dateisystem oder per Web erreichbare HTTP-Server und liest von dort Dokumente zur Verarbeitung ein (Schritt 1). Die Auswahl der Dateien lässt sich mit Hilfe einer Positiv- und einer Negativliste auf der Basis regulärer Ausdrücke sehr genau steuern.

Die Dokumente werden je nach Dateityp anhand der Dateiendung einem dafür zuständigen Präparator zugeführt, der die relevanten Inhalte extrahiert (Schritt 2). Diese können dann mit Hilfe von Lucene indexiert und für die spätere Nutzung dauerhaft hinterlegt werden (Schritt 3). Hierbei werden nicht die Dokumente an sich abgespeichert, sondern zu jeder Datei lediglich ein Eintrag mit wichtigen Nutzinformationen. Diese sind:

- Technische Angaben zum Originaldokument wie die Dateigröße, der Zeitpunkt der letzten Änderung und ein Verweis (URL)
- Im Dokument vorhandene Metadaten wie der Autor oder eine kurze Zusammenfassung
- Die eigentlichen Schlagworte aus dem rein textlichen Inhalt

Das Such-Modul von Regain schließlich greift bei einer Anfrage asynchron auf die zuvor hinterlegten Indexe zu.

Für die Belange von IRIS reicht die Funktionalität jedoch nicht aus. Folgende Aspekte gilt es zu berücksichtigen:

- Der geschriebene Lucene-Index ist gewissermaßen „starr“. Aufgrund seiner für das Lesen optimierten Datenstruktur lassen sich Einträge nur neu hinzufügen oder löschen (analog SQL-Anweisungen NEW bzw. DELETE); eine Funktion zum direkten Abändern eines Eintrages (analog einer SQL-UPDATE-Anweisung) existiert nicht. In IRIS müssen gewisse Attribute wie etwa die Zugehörigkeit eines Dokumentes zu einer Kategorie jedoch veränderlich sein.
- Die Datenquelle eines Projektpartners kann eventuell nur für einen bestimmten Zeitraum zum Indexieren zur Verfügung stehen.

- Neben lokalen Dateien und dem Web sollen auch Datenbanken als Datenquelle dienen können.
- Projektpartner sollen angebunden werden können, indem sie selbst qualifizierte Beschreibungen von Dokumenten generieren und diese in einer XML-Metadaten-Datei zu Verfügung stellen (insbesondere für kostenpflichtige Dokumente).
- Die textlichen Inhalte der erfassten Dokumente sollen für Auswertungszwecke (Begriffe-Netz bzw. Lernendes System) verfügbar sein. Bei Lucene wäre dies grundsätzlich durch das Setzen eines Schalters für das betreffende Feld (Speicher-Strategie „Store“) möglich, wodurch der Textinhalt zusätzlich unverändert als Ganzes beim Eintrag mit abgelegt wird. Das würde jedoch die Größe des Index nahezu verdoppeln und zudem dazu führen, dass dieser Textinhalt beim Auslesen des Index bei jeder neuen Suche (!) für jeden Treffer (!) in den Speicher geladen wird. Beides führt zu einer deutlichen Verschlechterung der Performance des Systems. Das ist unbedingt zu vermeiden!

Um die besonderen Belange von IRIS umzusetzen wurden folgende Modifikationen an „Regain“ vorgenommen:

- Hinzufügen eines Moduls zum Auslesen von Datenbank-Einträgen als „Dokumente“ (Pseudo-Dokumente genannt) und entsprechender Behandlung bei der Verarbeitung (neuer Präparator „DbPreparator“). Hierfür waren zusätzlich Änderungen am eigentlichen Crawler-Modul und der internen Dokumentenverarbeitung notwendig.
- Hinzufügen eines Moduls zum Auslesen von XML-Metadaten-Dateien mit Angaben über Dokumente. Damit kann ein Kooperationspartner beispielsweise geschützte, kostenpflichtige Dokumente von IRIS erfassen lassen, indem er selbst aus seinem Datenbestand eine XML-Datei generiert, in der für jedes Dokument etwa eine kurze Zusammenfassung, wichtige Schlagwörter, die Größe und – ganz wichtig – die URL zum Dokument hinterlegt werden. Diese Angaben kann IRIS technisch gesehen wie Datenbank-Einträge als Pseudo-Dokumente verarbeiten, ohne Zugriff auf die eigentlichen Dokumente haben zu müssen. Die Kontrolle, welche Daten zur Verfügung gestellt werden, sowie die Umsetzung eines Zugriffsberechtigungs-systems und/oder eines Bezahl-systems verbleiben in der Verantwortung des Projektpartners.
- Die Angaben zu den in IRIS erfassten Dokumenten werden zunächst nicht direkt in einem Lucene-Index erfasst, sondern in der eigenen Datenbank. Dafür wurde ein Modul zum Schreiben von Dokument-Einträgen analog dem „Regain“-Modul zum Schreiben von Lucene-Einträgen erstellt. Dadurch können die Attribute mit den üblichen Datenbank-Methoden (SQL) ausgelesen und – soweit sinnvoll – von der IRIS-Oberfläche (Administrator- bzw. Redaktions-Bereich) aus bearbeitet werden.

- Die Arbeitsweise des Crawlings in IRIS ändert sich aufgrund der Modifikationen und zerfällt in zwei Schritte, dargestellt in Abbildung 26.

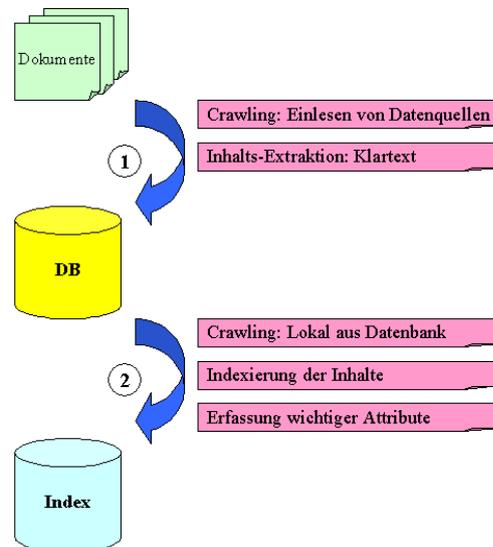


Abbildung 26: Schritte beim Crawling in IRIS

Im ersten Schritt werden die zu erfassenden Dokumente, Webseiten oder Datenbank-Inhalte systematisch nacheinander eingelesen. Bei jedem Dokument werden der Inhalt und gegebenenfalls Meta-Daten mit Hilfe eines Präparators extrahiert, d. h. reiner Text ohne Formatierungszeichen, Grafiken usw. Anschließend wird für jedes Dokument ein Eintrag mit allen wesentlichen Angaben in der lokalen IRIS-Datenbank erstellt (bzw. aktualisiert). Dies sind mindestens die URL, die Original-Größe und der Text-Inhalt. Falls vorhanden weiterhin das Datum der letzten Änderung des Dokuments, der Autor, der Titel, eine kurze Zusammenfassung und Überschriften bzw. Schlagzeilen.

Die folgende Abbildung 27 veranschaulicht den Weg, den jedes zu erfassende Dokument von der Quelle bis zum formatbereinigten Eintrag in der IRIS-Datenbank durchläuft.

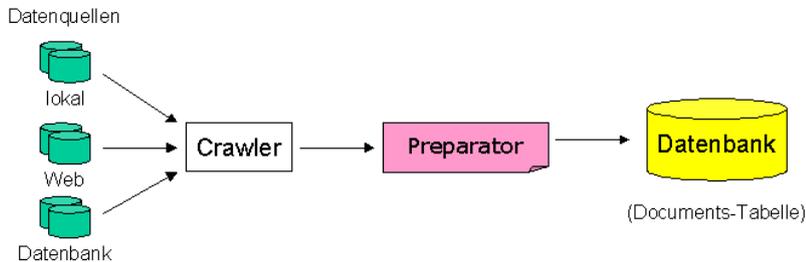


Abbildung 27: Crawling-Schritt 1: Einlesen der Dokumente

Der zweite Schritt erfolgt rein lokal. Die in der Datenbank abgelegten formatbereinigten Dokumenten-Abbilder werden mit Hilfe des Datenbank-Moduls ausgelesen. Abbildung 28 zeigt den Verarbeitungsweg der Dokumente von der Datenbank hin zum Index. Zu beachten ist die unterschiedliche Darstellung der Dokumenten-Tabelle der IRIS-Datenbank in den beiden Abbildungen, die aus der unterschiedlichen Funktion resultiert, nämlich einmal als Ziel (Abbildung 27) und einmal als Datenquelle (Abbildung 28).

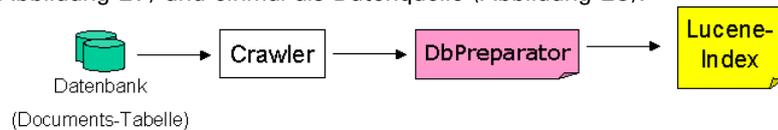


Abbildung 28: Crawling-Schritt 2: Indexierung der Inhalte

Die Zwischenspeicherung von Dokumenten-Einträgen in der IRIS-Datenbank löst einige der weiter oben aufgeworfenen Probleme:

- Änderungen wie beispielsweise die Zugehörigkeit eines Dokumentes zu einer bestimmten Kategorie oder die Zugriffsberechtigung sind jederzeit möglich.
- Die erfassten Inhalte stehen für Auswertungen jederzeit zur Verfügung. Das Originaldokument lässt sich im Allgemeinen daraus dennoch nicht rekonstruieren.
- Die IRIS-Administration bzw. -Redaktion kann sich problemlos einen Überblick über die im System erfassten Dokumente verschaffen.
- Der Index wird nicht für Auswertungszwecke belastet.

Die Suche im Lucene-Index schließlich verweist dabei nicht mehr direkt auf ein Dokument, sondern auf den entsprechenden Eintrag dazu in der IRIS-Datenbank. An dieser Stelle setzt denn auch das Berechtigungssystem von IRIS an und entscheidet, ob der Nutzer Zugriff auf das tatsächliche Dokument bekommt oder nicht.

10.4.6 Suche

Wird eine Suche ausgeführt (im Beispiel nach dem Begriff „Holz“, Abbildung 29), kommt zunächst dieselbe Sprachverarbeitung wie beim Indexieren zum Zug und anschließend eine von den Einstellungen abhängige Umsetzung der Suchanfrage in eine für die Indexierungsbibliothek verständliche Form (Lucene Suchanfrage-Syntax). Im Beispiel sei die Microsoft Word Datei „/daten/beispiel.doc“ mit Hilfe des Crawlers von Regain in einem Index erfasst worden. Nun sollen etwa die Felder „title“ und „content“ nach Übereinstimmungen mit dem Term „holz“ durchsucht werden. Anschließend sind die Ergebnisse mit den Treffern, absteigend sortiert nach Relevanz, zu präsentieren. Hier kann nun beispielsweise die zuvor hinterlegte Zusammenfassung (Feld „summary“) gezeigt werden. Zur Berechnung der Relevanz wird derzeit der Standard-Ranking-Algorithmus von Lucene auf Basis des Termfrequenzansatzes benutzt.

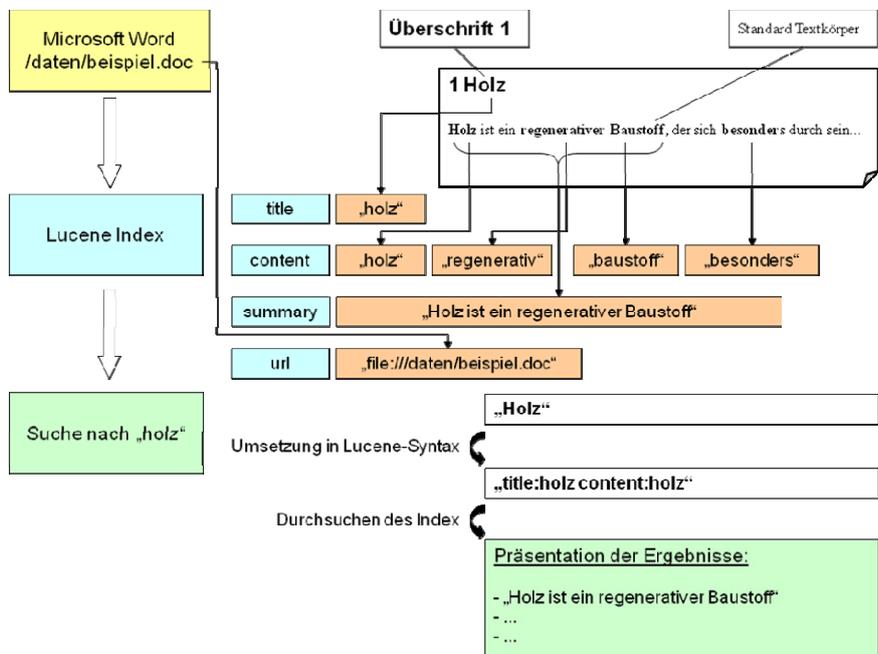


Abbildung 29: Beispiel für die Suche in Regain

Die einzelnen Aktionen sind in Abbildung 29 dargestellt. Aus der Datei „/daten/beispiel.doc“ (ausschnittsweise in der rechten oberen Ecke dargestellt) werden beim Indexierungsvorgang Inhalte extrahiert und in entsprechende Felder des Lucene-Index geschrieben (exemplarisch aufgezeigt wird die Befüllung der Felder „title“, „content“, „summary“ und „url“). Die Schritte bei der Suche finden sich in der unteren Bildhälfte. Zunächst wird die Suchanfrage („Holz“) in die Lucene-Syntax umgesetzt (hier sollen die Felder „title“ und „content“ nach dem Term „holz“ durchsucht werden). Alle Einträge, bei denen – wie in diesem Beispiel – wenigstens ein Feld die Suchkriterien erfüllt, werden in die Ergebnisliste aufgenommen. Für jeden solchen Treffer wird ein normierter Relevanzwert („Score“) berechnet, der u. a. von der Häufigkeit eines Suchterms in den Suchfeldern sowie im gesamten Dokumentenbestand abhängt. Die Score-Werte zweier Suchanfragen sind nicht vergleichbar. Nach der Suche kann die Liste absteigend nach Relevanz sortiert dargestellt werden.

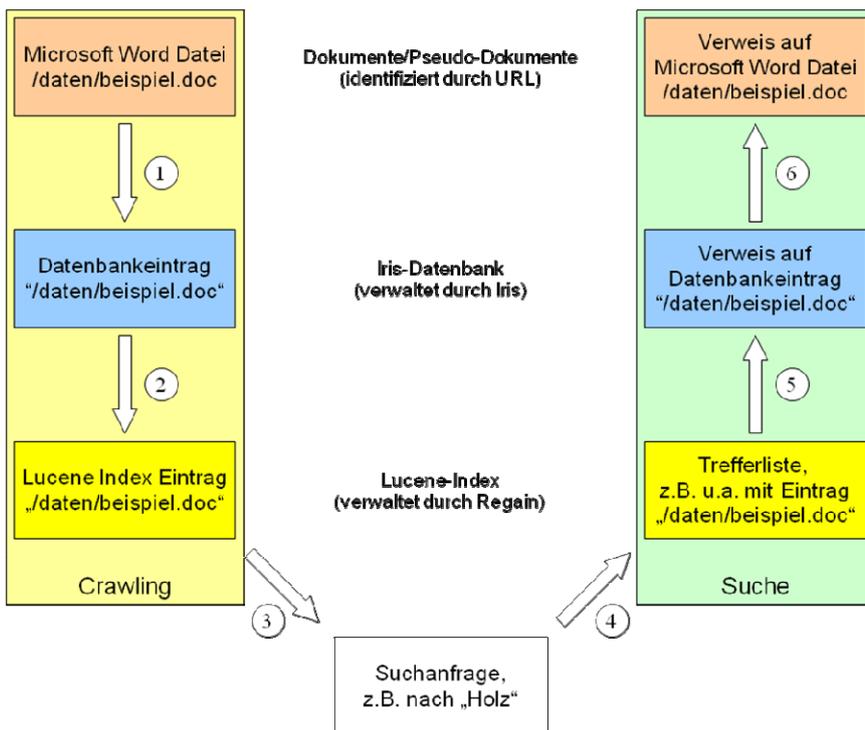


Abbildung 30: Beispiel für die Verarbeitung eines Dokuments in IRIS

Für IRIS wurde jedoch der Crawling-Prozess aufgrund verschiedener Erfordernisse in zwei Teilschritte aufgeteilt (vgl. Kapitel 10.4.5). Dementsprechend unterscheidet sich in IRIS auch der Suchprozess. Für den Nutzer ist dies transparent. Die Abbildung 30 zeigt zum Verständnis des modifizierten Crawling- wie Suchvorganges noch einmal den gesamten Verarbeitungsweg der Beispieldatei „daten/beispiel.doc“ auf. In Schritt 1 des Crawlings in IRIS wurde ein Datenbankeintrag geschrieben (1), in Schritt 2 der zugehörige Eintrag in einem Lucene-Index (2), sodass das System für die Suchanfrage bereit ist (3). Wird diese schließlich gestellt (4), veranlasst das System eine gewöhnliche Suche im Index, wie sie zuvor dargelegt wurde. Der einzige Unterschied ist, dass bei der Umsetzung der Anfrage in Lucene-Syntax die über Regain hinausgehenden Belange wie die mögliche Einschränkung auf bestimmte Kategorien wie etwa „Lexikon“ oder „Wissenschaftliche Artikel“ mit berücksichtigt werden. Die aus dem Index generierte Trefferliste kann dem Nutzer direkt gezeigt werden. Ein Klick auf einen Eintrag führt jedoch nicht wie bei Regain direkt zum zugehörigen Dokument, sondern stattdessen zunächst zum Datenbankeintrag von IRIS (5). Dadurch kann ein dynamisches Zugriffskontrollsystem sowie die Vorhaltung von Zusatzinformationen realisiert werden. Nur wenn der Benutzer aus Sicht von IRIS die notwendige Berechtigung besitzt, wird an das eigentliche Dokument, im Beispiel also die Datei „/daten/beispiel.doc“, weiterverwiesen und gegebenenfalls an den Nutzer übermittelt (6).

Es unterscheidet sich also streng genommen nicht der Suchvorgang als solcher, sondern die Interpretation der Suchergebnisse. Um ein Dokument vom Server abrufen zu können, muss es erstens als Eintrag im Index enthalten sein (eine obligatorische Sicherheitsprüfung von Regain) und zweitens muss der Nutzer das Recht dazu besitzen. Diese Überprüfung ist also zweistufig. Damit können dem Nutzer zumindest Angaben auch über solche Dokumente gezeigt werden, für die er kein Recht zum Abruf besitzt, etwa Zusatzinformationen zur eigentlichen Bezugsquelle. Mit dem in Regain integrierten Zugangskontrollsystem wäre dies so nicht möglich.

10.4.7 Verbindung zum Lexikon und dem semantischen Netz

Die in JSPWiki integrierte interne Suchfunktion auf den Wiki-Seiten („Finde im Lexikon“) schreibt für sich selbständig einen eigenen Lucene-Index mit allen notwendigen Angaben. Bezüglich der Suche im Gesamtsystem wird ausgenutzt, dass die Wiki-Seiten und deren Dateianhänge in der IRIS-Datenbank und nicht im Dateisystem abgelegt werden. Mit Hilfe der Datenbank-Crawling-Funktion werden die Artikel und Anhänge in regelmäßigen Abständen abgerufen. Die Lexikon-Einträge werden grundsätzlich wie alle anderen erfassten Dokumente behandelt. Sie werden einer eigenen Kategorie „Lexikon“ zugeordnet.

Das semantische Netz wird zunächst unabhängig von der Indexierung eingepflegt. Bei einer Suchanfrage wird das Netz darüber verständigt, sodass es die dazu passende Ontologie darstellen kann. Wird umgekehrt ein Begriff im semantischen Netz ausgewählt, wird dieser wie bei einer einfachen Suche verarbeitet (vgl. Kapitel 10.5).

10.4.8 Umsetzung

Mit „Regain“ ist eine Volltextindexierung mit einer einfachen Suche ähnlich wie bei „Google“ im Rahmen der angebotenen Funktionalität relativ leicht erreichbar. Für die Verwendung im Gesamtsystem IRIS musste es jedoch in einigen Teilen umgeschrieben werden. So ist in „Regain“ beispielsweise noch keine direkte Anbindung einer anderen Datenbank vorgesehen. Auch eine Auswertung von Metadaten aus XML-Dateien gibt es nicht. Andere Dinge mussten hinzugefügt werden wie z. B. Filter, um die Ergebnisse bei der Suche auf bestimmte Themengebiete einzuschränken, oder die Möglichkeit, sich beim Crawlen mit einem Benutzernamen und Passwort zu autorisieren. Der benötigte Arbeitsumfang für die Anpassungen überstieg spürbar die ersten Abschätzungen. Sie konnten aber dennoch innerhalb der Projektlaufzeit realisiert werden.

10.4.9 Resümee

Die Bibliothek Lucene bietet sich für den Einsatz in IRIS als das geeignete Instrument an. Der Funktionsumfang wie auch die Performance beim Schreiben und beim Lesen ist sehr zufrieden stellend. Den gesamten Index in einer Datenbank zu speichern, ist nicht notwendig und wohl auch gar nicht wünschenswert. Das „Lernen“ bzw. das „intelligente Verhalten“ als besonderer Vorzug von IRIS ist der nächste Schritt nach dem Indexieren. Der Index ist hierbei Grundlage und Datenbasis für weitergehende Verarbeitungen (z. B. statistische Auswertungen, eventuell in Kombination mit protokollierten Suchanfragen).

10.5 Semantisches Netz

Semantische Netze²⁹ dienen der Darstellung von Zusammenhängen und stellen so eine spezielle Art der Repräsentation von Wissen dar. Obwohl das grundlegende Konzept sich in einem formalen Modell ausdrückt, das Begriffe und Beziehungen umfasst, existieren dennoch verschiedene Varianten mit jeweils unterschiedlicher Ausprägung, Komplexität und Leistungsfähigkeit.

In semantischen Netzen werden Zusammenhänge prinzipiell als Aussagen in Form syntaktischer Tripel repräsentiert. Ein solches Tripel besteht aus einem Subjekt, einem Prädikat und einem Objekt (Abbildung 31).

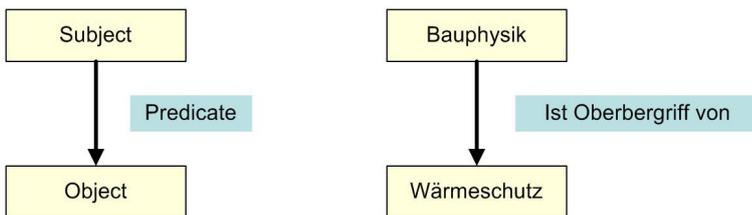


Abbildung 31: Das allgemeine Tripel und ein Beispiel für eine konkrete Aussage

Ein semantisches Netz, das eine beliebige Anzahl an Subjekten, Prädikaten und Objekten enthält, ist eine Sammlung syntaktischer Tripel. Der Vernetzungseffekt entsteht dadurch, dass jeder Begriff mit einer beliebigen Anzahl anderer Begriffe in Beziehung stehen kann. Ob ein Begriff Subjekt oder Objekt ist, hängt von der Richtung der Beziehung ab. Ein Begriff nimmt erst durch eine Beziehung eine Rolle an. Er kann also durchaus – bedingt durch seine Beziehungen – gleichzeitig Subjekt und Objekt sein.

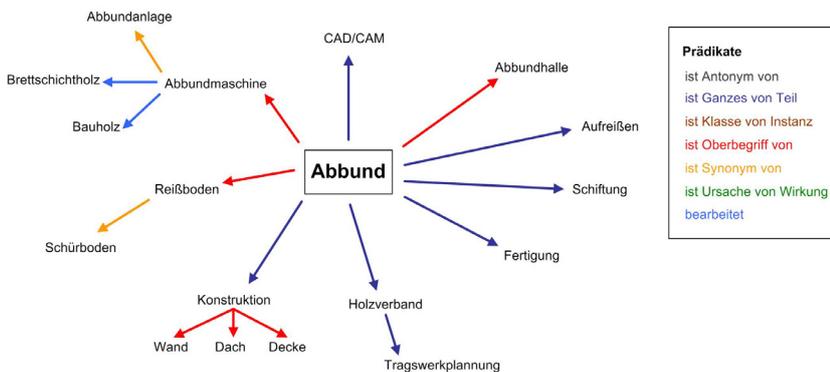


Abbildung 32: Beispiel für einen Auszug aus dem semantischen Netz

²⁹ Obwohl der Terminus „Semantische Netz“ treffend und korrekt ist, hat sich herausgestellt, dass er für den Anwender nicht intuitiv verständlich ist. Aus diesem Grund wurde in IRIS für den Anwender des Portals stattdessen der nicht ganz treffende Terminus „Begriffe-Netz“ verwendet.

10.5.1 Verwendung in IRIS

In IRIS unterstützt das semantische Netz den Benutzer insbesondere beim Suchvorgang. Es macht Zusammenhänge sichtbar und erweitert die herkömmlichen Suchmethoden (Volltext, Schlagwort usw.). Daraus ergeben sich unter anderem folgende Vorteile:

- Das Navigieren durch das semantische Netz hilft dem Benutzer, das richtige Vokabular zur Fragebildung zu finden und somit das passende Ergebnis effektiv und effizient zu erlangen
- Der Benutzer kann sich ein Themengebiet sowohl begrifflich als auch hinsichtlich der Zusammenhänge schnell erarbeiten. Möchte er einen Punkt vertiefen, so kann er unmittelbar in passenden Quellen nachlesen
- Eine komfortable visuelle Navigation per Mausklick schafft zudem einen guten Überblick und macht die Suche schnell

10.5.2 Software-Recherche

Prinzipiell gliedert sich das semantische Netz in IRIS in zwei Komponenten, eine für das Modell und eine für die Darstellung. Darüber hinaus wird für die dauerhafte Speicherung eine SQL-Datenbank benötigt, auf die vom Modell aus zugegriffen wird.

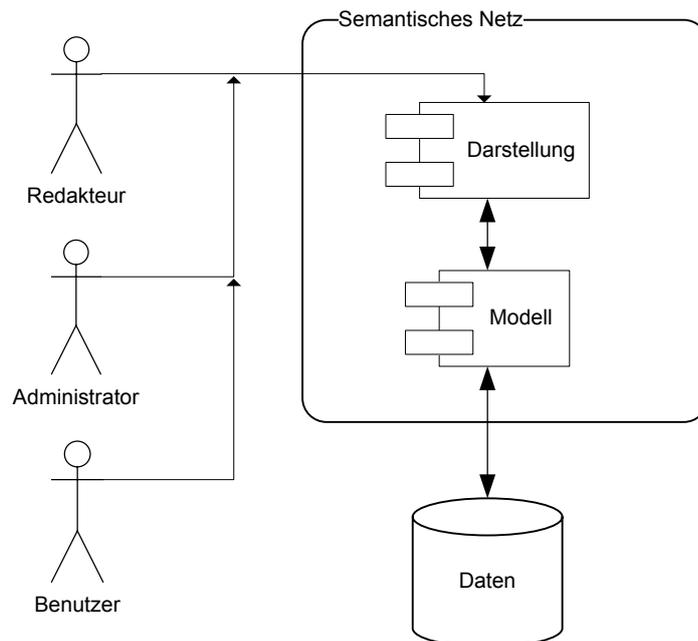


Abbildung 33: Struktur des semantischen Netzes

Die Recherche hat eine Reihe von als Open Source verfügbaren Werkzeugen als Basis für beide Komponenten ergeben. Darüber hinaus gibt es auch Systeme, die beide Komponenten beinhalten. Diese sind allerdings nur als kommerzielle Produkte erhältlich.

Für das Modell wurde aus der Menge verfügbarer Lösungen nach eingehender Überprüfung zunächst Jena 2.3 ausgewählt. Jena erfüllt folgende Anforderungen:

- Freie Verfügbarkeit (Open Source)
- Programmiersprache Java
- Realisiert die gängigen Standards
- Ausgereifte Software mit weitem Verbreitungsgrad
- Bietet zusätzliche Erweiterungen

Auf das Modell setzen die Darstellungen des semantischen Netzes auf, die sowohl für die Redaktion als auch für die Suche bzw. Navigation benötigt werden. Während für die Redaktion textliche Varianten zweckmäßig sind, die sich mit Standard-Mitteln realisieren lassen, ist bei der Darstellung für den Benutzer eine graphische Variante sinnvoll.

Für diese visualisierte Darstellung wurde zunächst mit einer Eigenentwicklung begonnen. Der Lösungsansatz genügte jedoch nicht den Erfordernissen und somit wurde das Open-Source-Produkt TouchGraph mit interessanten und vielversprechenden Eigenschaften untersucht. TouchGraph bietet insbesondere:

- Freie Verfügbarkeit (Open Source)
- Programmiersprache Java
- Interaktive, dynamische Darstellung der Knoten und Beziehungen
- Ermöglicht intuitives Handeln des Benutzers

10.5.3 Realisierung

Jedoch konnten weder Jena noch TouchGraph zu einer zufriedenstellenden Lösung beitragen. Die entscheidenden Probleme waren:

- Da TouchGraph als Java-Applet im Browser auf der Clientseite ausgeführt werden muss, muss auch Jena dort ausgeführt werden. Ein erheblicher Teil der Software und der damit auszuführenden Funktionalität wird somit auf den Client verlagert. Daraus entstehen spezielle und hohe Anforderungen an den Client, an die Kommunikation zwischen Client und Server und an die Programmierung
- Schwieriges Zusammenspiel der Komponenten aufgrund der deutlich unterschiedlichen Ansätze bei Jena und TouchGraph
- Die dynamische Datenhaltung der in Jena XML-basierten Daten in einer SQL-Datenbank wäre aufwändig zu realisieren

In Konsequenz wurde eine leistungsfähige Eigenentwicklung konzipiert und realisiert. Diese Lösung bietet folgende Vorteile:

- Modell und Darstellung passen nahtlos aneinander
- Die Modelloperationen und der Aufbau der graphischen Darstellungen werden ausschließlich auf der leistungsstarken Serverseite ausgeführt
- Sie kann von jedem gängigen Client problemlos verwendet und dargestellt werden, weil dort nur Standard-Funktionalität benötigt wird
- Es wird ein ebenfalls eigenentwickeltes spezialisiertes Modell eines semantischen Netzes verwendet, das einerseits nur sehr einfache Aussagen zu-

lässt und andererseits mit zusätzlichen Merkmalen angereichert ist, um eine praktikable Lernfähigkeit zu erlangen

10.6 Lexikon

10.6.1 Wiki

IRIS beinhaltet zur weiteren Informationsbereitstellung ein Lexikon mit Anhängen. Da davon ausgegangen werden musste, dass die Redakteure keine tieferen Informatik- und Programmierkenntnisse besitzen, musste Ihnen ein Instrument an die Hand gegeben werden, mit dem sie Inhalte in das Lexikon eingeben können, ohne dass sie HTML-Quellcode produzieren müssen. Eine Möglichkeit für ein solches Instrument ist ein so genanntes Wiki.

Ein Wiki ist eine zumeist im Internet verfügbare Seitensammlung bzw. das zum Betrieb (Einstellen, Bearbeiten und Löschen von Seiteninhalten) von eben diesen Seiten verwendete Programm. Der Name stammt von „wikiwiki“, dem hawaiianischen Wort für „schnell“.

Wikis bringen einige Eigenschaften bzw. Konzepte mit, die die Systematik von IRIS unterstützen. Wikis sind leichtgewichtige Tools. Sie bieten, trotz der im Vergleich zu den meisten Web-Content-Management-Systemen (WCMS) geringen Menge an Quellcode, eine umfassende Funktionalität. Der Inhalt von Wikis wird als evolutionäres Gebilde angesehen, das sich laufend im Wandel befindet und somit den Wissen schaffenden Prozess (lernendes System) unterstützt. Wikis stellen Kontexte durch Links her. Auf eine vom System vorgegebene hierarchische Ordnung und Metainformationen wird meist völlig verzichtet, im Vordergrund steht die Vernetzung. Die Artikel sind prinzipiell alle gleichwertig und bilden, untereinander verbunden über Hyperlinks, ein Netz von Informationen. Die geplante Loslösung von der klassischen Verschlagwortung und einer festen hierarchischen Struktur in IRIS kann so leicht umgesetzt werden.

Ein weiteres Grundkonzept von Wikis ist, dass jeder Besucher der Webseite jeden Artikel verändern darf. Dies ist für das IRIS-Lexikon jedoch nicht sinnvoll. Daher werden Redaktionsrechte über die für die Informationsplattform vorhandene Benutzerverwaltung eingeschränkt.

10.6.2 JSPWiki

Aus einer Menge von verfügbaren Wikis wurde nach eingehender Prüfung JSPWiki ausgesucht. Dieses Wiki erfüllt folgende an es gestellte Anforderungen:

- Freie Verfügbarkeit (Open Source).
- Programmiersprache Java
- Speicherung der Daten in einer MySQL-Datenbank
- Ausgereifte Software mit weitem Verbreitungsgrad
- Nützliche Plug-Ins verfügbar oder in Entwicklung

Natürlich bot auch JSPWiki nicht jede gewünschte Funktionalität des IRIS-Lexikons, deshalb waren einige Anpassungen nötig.

10.7 Lernen

10.7.1 Einführung

IRIS ging mit dem Anspruch an den Start, ein sich weiter entwickelndes und lernendes System zu werden. Um ein System mit Lernfähigkeit auszustatten, muss man eine Vorstellung davon haben, was Lernen im Allgemeinen und evtl. für ein bestimmtes System im Speziellen bedeutet.

Lernen heißt, etwas erkennen, von dem man davon ausgeht, dass es wesentlich, wichtig, richtig, wahr usw. ist. Es erfordert außerdem, das Erlernete zu merken, aber auch, es wieder – zumindest zum Teil – zu vergessen, wenn es nicht mehr die Eigenschaften besitzt, aufgrund derer es gelernt wurde.

Lernen bewirkt eine Veränderung, zunächst eine Veränderung des Wissens. Ein Computerprogramm muss dieses Wissen speichern und benötigt dazu eine geeignete Wissensrepräsentation. Wie auch immer diese Wissensrepräsentation gestaltet ist, es handelt sich stets um eine sogenannte Zustandsmaschine. Das konkrete komplette Wissen eines Systems entspricht einem bestimmten Zustand. Nach einem Lernvorgang liegt ein neuer Zustand vor. Lernen bedeutet also aus Sicht der Maschine immer eine Zustandsveränderung. (Die Umkehrung dieser Aussage gilt jedoch nicht.) Damit ein System für den Anwender sinnvoll ist, muss das darin enthaltene Wissen nach außen auch erkennbar und wirksam werden. Lernen bewirkt also hier auch eine Veränderung des Systemverhaltens.

10.7.2 Grundüberlegungen für IRIS

Es gibt eine Vielzahl von Ansätzen, computerbasierte Systeme mit einem gewissen Maß an Intelligenz und Lernfähigkeit auszustatten:

- Ausgehend von den natürlichen Lebensformen, sind es die neuronalen Netze. Diese eignen sich besonders für Aufgaben, bei denen ein explizites systematisches Wissen gering ist bzw. sich nur schwer formulieren lässt.
- Demgegenüber stehen streng formalisierte Methoden, insbesondere die Prädikatenlogik.
- Einen aktuellen Ansatz bilden semantische Netze, die im Falle einer ausgeklügelten Ausgestaltung (Aussagen über Aussagen, spezielle Ontologie usw.) umfangreiche Formulierungen und ein Schlussfolgern zulassen.
- Sogenannte Expertensysteme verwenden hauptsächlich logikbasierte Ansätze, sind aber durch ihre starke Spezialisierung in ihrer Anwendbarkeit praktisch immer auf ein Fachgebiet, meist sogar nur auf eine ganz bestimmte Aufgabe, beschränkt.

Während eine jeweils gewählte Methode den formalen Rahmen der Wissensrepräsentation bereits klar absteckt, so bleibt trotzdem ein erheblicher Spielraum bei der Bildung des konkreten Modells für die gegebene Aufgabe. Die Qualität des Modells entscheidet maßgeblich über den Erfolg im tatsächlichen Einsatz. Aber selbst wenn das Modell passt, so bleibt in der Regel ein entscheidendes Problem: Das System muss praktisch anwendbar sein und muss es auch bleiben, wenn sich die darin enthaltenen Informationen im Laufe der Zeit deutlich in Form, Größe und Inhalt ändern.

Für IRIS in seiner Gesamtheit galt es, von Anfang an den kritischen Punkt des Lernens im Auge zu behalten. Hätte man hier einen falschen Weg eingeschlagen, wäre dieser nicht mehr im Rahmen des Projektes korrigierbar gewesen. Nachfolgend kann nur ein grober Abriss für die vielen Entscheidungen auf dem Weg zu einer geeigneten Lösung gegeben werden.

Bestimmte Lösungsmöglichkeiten wurden verworfen:

Neuronale Netze sind für IRIS aufgrund folgender Punkte kaum geeignet:

- Die Aufgabe passt nicht zum möglichen Einsatzspektrum
- Die Modellierung ist problematisch
- Die vorhandene bzw. zu erwartende Datenmenge ist zu groß
- Der Inhalt ist zu dynamisch
- Die Steuerungsmöglichkeiten sind zu gering
- Das Lernergebnis kann nicht angepasst werden

Auch logikbasierte Verfahren sind nicht geeignet:

- Sie setzen ein exaktes Ausgangsmaterial und eine exakte Vorgehensweise bei der Auswertung voraus
- Daraus resultiert ein vollkommen unverhältnismäßig großer Aufwand, sowohl bei der Aufbereitung des Datenmaterials als auch bei der Auswertung

10.7.3 Realisierung

In IRIS findet Lernen in der Indexierung und im semantischen Netz statt.

Lernen durch Indexierung

Dadurch, dass IRIS als zentralisierende Plattform sowohl über die Indexierung als auch über das Lexikon auf externe Daten zugreift, verändert sich das über IRIS verfügbare Wissen auch dann, wenn das Datenmaterial in IRIS selbst nicht geändert wird. Das Lernen aus externen wie auch aus internen Informationen findet dann statt, wenn der Index neu aufgebaut wird, was durch eine entsprechende Konfiguration regelmäßig und automatisiert erfolgen kann. Die weiter oben beschriebene lernbedingte Zustandsveränderung schlägt sich im Index nieder, der als Metainformation für die Suche zur Verfügung steht.

Lernen im semantischen Netz

Beim semantischen Netz kommt ein statistisches Verfahren zum Einsatz. Hierzu wird zunächst für jeden Begriff bzw. für jede Beziehung die Zahl der Treffer bei der Suche und Navigation gezählt. Nach einer gewissen Reifungszeit kann man eine bestimmte Trefferzahl je Zeit feststellen. Hieraus wird über alle reifen Begriffe bzw. Beziehungen eine relative Kurzzeit-Relevanz berechnet. Die jeweils vorherige Kurzzeit-Relevanz wird in die Berechnung mit einbezogen. Aus dieser Kurzzeit-Relevanz und der weiteren Langzeitrelevanz wird eine neue Langzeit-Relevanz berechnet. Diese Langzeit-Relevanz wird bei der Navigation im semantischen Netz visualisiert bzw. textlich angezeigt. Weiterhin können durch Angabe eines Schwellwertes Begriffe bzw. Beziehungen, die diesen Schwellwert unterschreiten, bei der Anzeige unterdrückt werden.

Das mehrstufige Relevanz-Modell dämpft kurzfristige Schwankungen. Dadurch werden Verzerrungen vermieden, die z. B. aufgrund eines bestimmten aktuellen Themas durch stark vermehrte Suche in IRIS auftreten können. Da manche Begriffe und Beziehungen in ihrer Relevanz, z. B. aufgrund ihrer fachspezifischen Bedeutung, nicht verändert werden sollen, kann die Veränderbarkeit individuell abgeschaltet werden. Falls eine Variabilität nur innerhalb eines reduzierten Relevanz-Bereiches gefordert ist, so kann ein unterer und ein oberer Grenzwert entsprechend gesetzt werden.

Auch neue Begriffe und Beziehungen können gelernt werden. Ist ein gesuchter Begriff im semantischen Netz noch nicht vorhanden, so wird dieser angelegt. Eine neue Beziehung wird dann eingetragen, wenn zwischen zwei unmittelbar nacheinander gesuchten Begriffen noch keine Beziehung bestand. In beiden Fällen erfolgt der jeweilige Eintrag mit dem Attribut „ungültig“ und einem negativen Zählerwert. Damit ein Eintrag gelernt wird, muss die Zählerwert-Schwelle von Null überschritten werden. Andernfalls wird der Eintrag beim Lernvorgang wieder entfernt. Man kann das mit dem Informationsübergang vom Kurzzeit- ins Langzeitgedächtnis vergleichen.

Wesentliche Eigenschaften des Lernens in IRIS sind:

- Die verwendeten Verfahren sind in der Anwendung unkompliziert, verhalten sich stabil und tolerant
- Der Lernvorgang kann einfach angestoßen werden und erfolgt bei der Indexierung in der Regel automatisch
- Weiteres Datenmaterial kann problemlos hinzugefügt werden ohne die Lernfähigkeit dadurch zu stören
- Das Lernergebnis kann beeinflusst werden
- Es gibt keine Metadaten, die der Administrator oder der Redakteur bearbeiten oder verwalten müsste
- Es ist keine spezielle Anpassung des Verfahrens für den Holzbau erforderlich

10.7.4 Ausblick

Manueller vs. automatischer Start des Lernvorganges

Der Indexierungsvorgang kann wahlweise automatisch gestartet werden. Im Prinzip könnte das auch für das semantische Netz realisiert werden. Da hier aber für das konkrete Lernverhalten noch keine ausreichenden Erfahrungen vorliegen, lässt sich der Lernvorgang im semantischen Netz vorerst nur manuell starten. Zudem ist es durchaus sinnvoll, wenn der redaktionelle Experte sein Wissen einfließen lässt, indem er gegebenenfalls die Attribute der Einträge anpasst. Zudem sollte eine neue Beziehung mit einem treffenden Prädikat ausgestattet werden, das der Redakteur vorgeben kann.

Lernen und Vergessen

Bei einem Index ist es natürlich naheliegend, dass auf Dinge, die nicht mehr vorhanden sind, auch nicht mehr verwiesen wird. Wenn es sich dennoch um wichtige Informationen handelt, sollte das Archivieren eher durch einen Redakteur geschehen, als dem System eine entsprechende Intelligenz abzuverlangen.

Beim semantischen Netz, das ja mit Relevanzen arbeitet, die sich aus dem Suchverhalten des Benutzer ergeben, könnte es durchaus sinnvoll sein, das momentan realisierte Verhalten zu verändern. Evtl. orientiert man sich hierfür am „natürlichen“ Lernen des Menschen. Hier könnte man sich beispielsweise vorstellen, dass das Vergessen langsamer verläuft als das Lernen.

11 IRIS-Software im Überblick

Nachfolgend wird ein Überblick über die wichtigsten Funktionen des konkreten IRIS-Software-Systems gegeben³⁰. Da IRIS über das Internet verfügbar ist, kann das hier Dargestellte unter heute üblichen Voraussetzungen ohne weiteres nachvollzogen und im Detail vertieft werden. Konkret sind ein Computer mit Zugang zum Internet und ein auf diesem Computer installierter gängiger Webbrowser, kurz Browser genannt, erforderlich.

Derzeit ist IRIS unter der Adresse <http://iris.fh-rosenheim.de/> erreichbar. Falls sich die Adresse nach Projektabschluss geändert hat, kann die neue Adresse an der Hochschule Rosenheim, an der TU München, bei der DGfH bzw. bei den Entwicklern nachgefragt werden.

11.1 Erläuterungen zur Verwendung von IRIS

Starten von IRIS

IRIS wurde als sogenannte Web-Applikation realisiert, die auf einem einzelnen Server betrieben wird. IRIS wird durch die Administratoren gestartet und läuft von da an für eine möglichst lange Zeit durch. Der Programmablauf von IRIS wird normalerweise nur zu Wartungsarbeiten unterbrochen. Obwohl IRIS nur eine sogenannte Programm-Instanz verwendet, kann zu jedem Zeitpunkt eine größere Anzahl von Anwendern das System gleichzeitig benutzen, ohne dass die einzelnen Anwender voneinander etwas merken. Hierfür wird für jeden Anwender, der an Nutzung von IRIS teilnimmt, eine sogenannte Sitzung³¹ angelegt und verwendet.

Ein Anwender, der an der Nutzung von IRIS teilnehmen möchte, muss lediglich in dem von ihm verwendeten Browser o. g. Adresse eingeben, um auf die Startseite von IRIS zu gelangen. Der Start eines speziellen IRIS-Programmes ist weder erforderlich noch möglich.

Beenden von IRIS

Um die Teilnahme an der Nutzung von IRIS zu beenden, muss der Anwender nichts Spezielles unternehmen. Falls der Anwender angemeldet ist, empfiehlt es sich jedoch, dass er sich abmeldet. Andernfalls bleibt seine Sitzung noch für eine gewisse Zeit am Server gültig. Ansonsten kann der Browser problemlos geschlossen werden.

³⁰ Dieser Überblick ist nicht als Anleitung zu verstehen, weder für die Benutzung noch für die Verwaltung des Systems.

³¹ Wichtig hierfür ist, dass im Browser die Cookies-Funktion aktiviert ist, was standardmäßig der Fall ist.

11.2 Gliederung der Anwendung

11.2.1 Gestaltung und Aufbau der Oberfläche

Alle Seiten in IRIS haben ein durchgängiges Erscheinungsbild. Es gibt zwei prinzipielle Darstellungsvarianten. Das sind einerseits die normale Darstellung für die eigentliche Verwendung von IRIS und andererseits Pop-up-Fenster, mit denen spezielle Informationen – insbesondere Hilfetexte – ausgegeben werden, ohne dass dadurch die normale Darstellung verändert wird.

Der Aufbau der normalen Darstellung wird anhand der Startseite (Abbildung 34) erklärt:

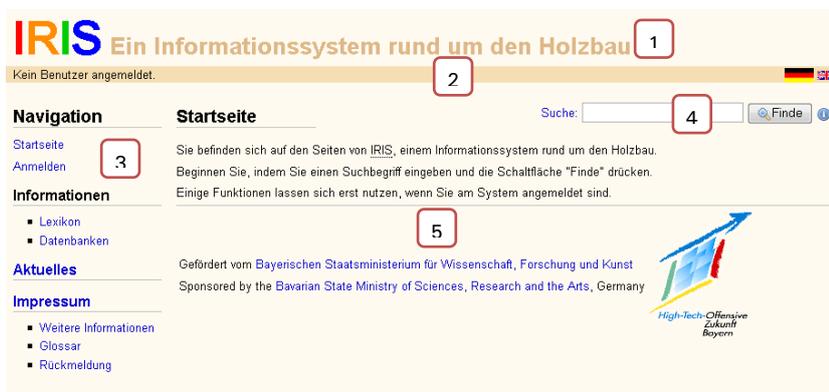


Abbildung 34: Die Startseite von IRIS

Die Seite ist folgendermaßen aufgeteilt:

- Die Titelzeile (1)
hat einen ausschließlich gestalterischen Zweck
- die Statuszeile (2)
stellt auf der linken Seite den angemeldeten Anwender namentlich dar und zeigt auf der rechten Seite die aktive und die alternative Dialogsprache an. Die Sprache kann – mit Ausnahme des Lexikons – an jeder Stelle durch Klicken auf das entsprechende Flaggensymbol gewechselt werden
- Über das Navigationsmenü (3)
können die einzelnen Programm-Bereiche von IRIS aufgerufen werden. Der Umfang der tatsächlich aufgeführten Menüpunkte hängt vom Anwenderstatus ab
- Über die Sucheingabe (4)
können alle indixierten und im semantischen Netz vorhandenen Informationen abgefragt werden
- Der Anzeige- und Dialogbereich (5)
stellt je nach aufgerufener Programmfunktion den eigentlichen Inhalt bzw. die Suchergebnisse dar oder dient zur Dateneingabe

Ein Pop-up-Fenster, das sich immer vor dem IRIS-Anwendungsfenster positioniert, hat keine spezielle Aufteilung wie Abbildung 35 exemplarisch anhand des IRIS-Glossars zeigt.

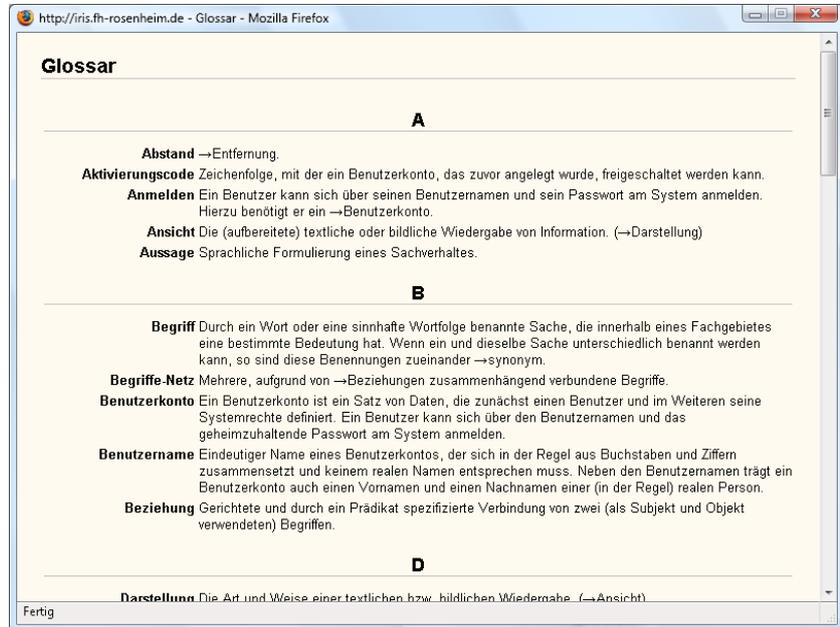


Abbildung 35: Das IRIS-Glossar als Beispiel für ein Pop-up-Fenster

11.2.2 Navigation

Es existieren folgende Möglichkeiten für das Navigieren:

- Navigationsmenü
- Suche
- Dialogabhängige Navigation

Ein Großteil der Navigation wird über das Navigationsmenü erledigt. Da es immer angezeigt wird, kann man sehr schnell von jeder Stelle aus eine andere Programmfunktion erreichen. Wie im HTML-Kontext üblich, reicht es aus, auf den jeweiligen textlich dargestellten Hyperlink mit der linken Maustaste zu klicken. Beispielsweise wird das vorstehende Glossar angezeigt, wenn man im Navigationsmenü den entsprechend benannten Link wählt.

Da die Suche einen wesentlichen Bestandteil von IRIS darstellt, ist auch diese jederzeit verfügbar.

Darüber hinaus gibt es aber auch eine Reihe von Funktionen, die nur von bestimmten Programmstellen aus aufgerufen werden können. Die Links befinden sich dann im Anzeige- und Dialogbereich.

Bei der Darstellung eines gewählten Links gibt es diese Alternativen:

- im selben Fenster bzw. im selben Tab
- in einem (neuen) Pop-up-Fenster
- in einem neuen Fenster bzw. einem neuen Tab

Während bei der normalen Darstellung innerhalb von IRIS die nächste Ansicht die aktuelle Ansicht ersetzt, wäre dieses Verhalten bei Links auf Seiten außerhalb von IRIS sehr ungünstig. Deshalb werden Links auf Seiten außerhalb von IRIS in einem neuen Browserfenster bzw. einem neuen Browsertabs (abhängig vom Browser und von der Browserkonfiguration) geöffnet.

11.2.3 Hilfe

Manche Funktionen sind nicht ohne weiteres anwendbar bzw. bieten mehr Funktionalität als unmittelbar ersichtlich ist. In diesen Fällen wird eine spezifische Hilfe angeboten, die über einen Link als Pop-up-Fenster aufgerufen werden kann. Ein einheitliches Symbol kennzeichnet den entsprechenden Link. An einigen Stellen kann die Hilfe auch über die zugehörige Beschriftung bzw. Überschrift aufgerufen werden. Beispielsweise ist der Link für die kontextspezifische Hilfe beim Suchdialog sowohl beim *i*-Symbol als auch bei der Beschriftung „Suche“ hinterlegt (Abbildung 36).



Abbildung 36: Beispiel für Stellen zum Aufruf kontextspezifischer Hilfe

11.3 Anmeldung am System

IRIS sieht eine Registrierung und Anmeldung am System vor. IRIS lässt sich aber auch ohne Registrierung und Anmeldung ohne weiteres sofort nutzen. Eine erweiterte Funktionalität ist allerdings erst dann möglich, nachdem man sich angemeldet hat (Abbildung 37).

Anmelden

Benutzername

Passwort

Hinweis

- Wenn Sie bereits registriert sind, melden Sie sich bitte mit Ihrem Benutzernamen und Ihrem Passwort an.
- Sollten Sie noch keinen Benutzerzugang haben, so registrieren Sie sich bitte. Die Registrierung ist kostenlos!

Weitere Funktionen

- [Neuen Benutzer anlegen](#)
- [Benutzerzugang freischalten](#)

Abbildung 37: Login-Dialog

Um sich anmelden zu können, benötigt man ein aktiviertes Benutzerkonto. Dieses erhält man, wenn man sich zunächst registriert (Abbildung 38) und das erzeugte Benutzerkonto anschließend frei schaltet. Alternativ kann ein Benutzerkonto vom Administrator angelegt werden.

Benutzer anlegen

Benutzername *	<input type="text"/>
Passwort *	<input type="password"/>
Passwort *	<input type="password"/> (wiederholt)
Anrede *	Bitte wählen... ▾
Titel	<input type="text"/>
Vorname *	<input type="text"/>
Nachname *	<input type="text"/>
Adresszusatz	<input type="text"/>
Straße und Hausnummer *	<input type="text"/>
PLZ, Ort *	<input type="text"/> <input type="text"/>
Land *	Bitte wählen... ▾
Telefonnummer	<input type="text"/>
Weitere Telefonnummer	<input type="text"/>
Emailadresse *	<input type="text"/>
Berufsgruppe	Bitte wählen... ▾
Sprache *	Bitte wählen... ▾
<input type="checkbox"/> Ja, ich stimme den Nutzungsbedingungen von IRIS zu.	
<input type="button" value="✓ OK"/> <input type="button" value="✗ Abbrechen"/> <input type="button" value="↺ Rücksetzen"/>	

Hinweis

- Die mit einem * markierten Einträge sind Pflichtfelder.

Abbildung 38: Dialog zum Anlegen eines Benutzers

11.4 Suche

11.4.1 Eingabe der Suche

In IRIS ist die Suche eine zentrale Funktion. Da IRIS eine Internet-Anwendung ist, kommt es besonders darauf an, dass jeder Anwender sofort eine Suche starten kann. Eine Anwendung wie IRIS, die für ein breites Anwenderspektrum geeignet sein soll, steht vor einer besonderen Herausforderung: Einerseits sollen die Funktionen ohne Einarbeitung intuitiv und einfach nutzbar sein. Andererseits sollen auch die Anforderungen anspruchsvoller Anwender erfüllt werden können.

Für die Suche steht in IRIS nur ein einzelnes Eingabefeld zur Verfügung.



Abbildung 39: Beispielhafte Suche nach dem Wort Holzbau

Ein naheliegender und äußerst praktischer Grund hierfür ist, dass der Anwender sich keine Gedanken darüber machen muss, in welche Felder er welche Suchinformation einzugeben hat. Demgegenüber strukturieren verschiedene andere Anwendungen die Suche hinsichtlich der Eingabefelder sehr stark. Meistens liegt dort der Grund in der Verwendung gleichermaßen strukturierter Metadaten.

Obwohl die Reduzierung auf ein Eingabefeld sich auf den ersten Blick wie eine funktionale Einschränkung darstellt, ist das dahinterliegende und zunächst nicht offensichtliche Konzept im Prinzip sogar mächtiger. Bei Anwendungen mit mehreren Eingabefeldern kommt jedem Feld eine bestimmte Funktion zu. Z. B. dient ein Feld zur Eingabe des Autors ein anderes zur Eingabe des Titels. Zudem ist meist festgelegt, ob alle Suchbedingungen gleichzeitig gegeben sein müssen oder ob nur eine der angegebenen Bedingung erfüllt werden muss.

Bei der in IRIS realisierten Lösung mit nur einem Eingabefeld unter Verwendung des Open-Source-Frameworks „Lucene“ gibt es verschiedene Möglichkeiten, eine komplexe Suche durchzuführen³²:

- Generell besteht der Text einer Suchanfrage aus Elementen, die beliebig kombiniert werden können. Setzt sich die Suchanfrage aus mehreren Elementen zusammen, dann dient das Leerzeichen zur Trennung einzelner Elemente voneinander
- Einfacher Term:
Suche nach Dokumenten, die diesen Term enthalten
- Phrase
Suche in Dokumenten nach der gesamten Phrase zwischen den Anführungszeichen
- Boolesche Operatoren
Terme und Phrasen können logisch mit den booleschen Operatoren UND (reserviertes Wort "AND" oder "&&"), ODER ("OR", "||") und NICHT ("NOT", "!") verknüpft werden. Alternativ können als Kurzschreibweise die Präfixe "+" (Pluszeichen) für UND bzw. "-" (Minuszeichen) für NICHT direkt vor einen Term oder eine Phrase gesetzt werden
- Suche mit Platzhalter
In einem Such-Term können Platzhalter verwendet werden. Dazu stehen das Fragezeichen ("?") für ein einzelnes Zeichen sowie der Stern ("*") für beliebig lange Zeichenketten (inkl. leerer Zeichenkette) zur Verfügung

Darüber hinaus gibt es noch weitere Möglichkeiten, die aber eher für forcierte Experten gedacht sind und deshalb in der IRIS-Hilfe nicht erläutert werden:

- Gruppierung von Elementen
durch Klammerung der booleschen Operationen
- Undeutliche Suche
- Distanzsuche
mit Angabe eines bestimmten maximalen Abstands für die gesuchten Wörter
- Suche in bestimmten Indexfeldern
Hier kann unterschieden werden, in welchem strukturellen Kontext das Gesuchte steht (insbesondere Inhalt, Titel, Überschrift, Dokument-Größe)
- Verstärkungsfaktoren
zur Gewichtung der gesuchten Wörter

Somit kann durch die Eingabe bestimmt werden, ob man eine einfache Suche oder eine mehr oder weniger differenzierte Suche durchführen möchte. Zudem ist ein Anwender, der einfach nur nach einen bestimmten Begriff suchen möchte, sofort im Stande dies auch auf einfachste Weise zu tun.

³² Da viele dieser Einstellungen nur den Index betreffen, werden diese für die Suche im semantischen Netz herausgefiltert. Dies geschieht intern, so dass der Anwender davon nichts mitbekommt.

Das Suchergebnis im Index wird nach der Treffergüte sortiert angezeigt. In der Auflistung werden neben der Treffergüte der Titel, ein textlicher Auszug und die Quelle des Dokumentes ausgegeben. Weitere Details erhält man, wenn man auf den gewünschten Listeneintrag klickt (Abbildung 43).

Details zum gewählten Suchergebnis Suche:

Details zum Dokument	
Link zum Dokument:	Holzbau der Zukunft
Original Adresse:	/iiswoodwiki/Wiki.jsp?page=Holzbau der Zukunft
Besitzer:	/opt/iisy/iisycrawlerconfig/WikiCrawling.xml
Letzte Änderung:	2007-10-23 10:47
Dateigröße:	8628
Kategorien:	Lexikon
Themengebiete:	Holz, Holzbau, Holzlexikon
Kurze Zusammenfassung:	Inhaltsverzeichnis Verbundforschungsprojekt - Holzbau der Zukunft (HdZ) Forschungsprojekte Weiterführende Informationen Siehe auch Weblinks Kategorien Verbundforschungsprojekt - Holzbau der Zukunft...

Abbildung 43: Detailbeispiel beim Index-Suchergebnis

11.4.3 Such- und Darstellungsoptionen

Standardmäßig umfasst das Ergebnis das semantische Netz und den Index. Beide Teilergebnisse sind in der Regel jedoch so umfangreich, dass sich das Gesamtergebnis auf dem zur Verfügung stehenden Platz nicht zufriedenstellend anzeigen lässt. Das trifft insbesondere auch dann zu, wenn der Anwender nur eine Anzeige mit geringer Auflösung hat. Deshalb gibt es direkt über dem Ergebnis die Möglichkeit, die Ansicht umzuschalten (Abbildung 44).

[Ansicht: **Alles** [Begriffe-Netz](#) [Ergebnisliste](#)] [[Einstellungen](#)] [Suche nach **'Holzbau'**]

Abbildung 44: Navigation der Such- und Darstellungsoptionen

Wenn nur das Teilergebnis des Begriffe-Netzes ausgewählt wird, dann besteht bei der graphischen Darstellung die Möglichkeit, auch weiter entfernte Begriffe anzuzeigen.

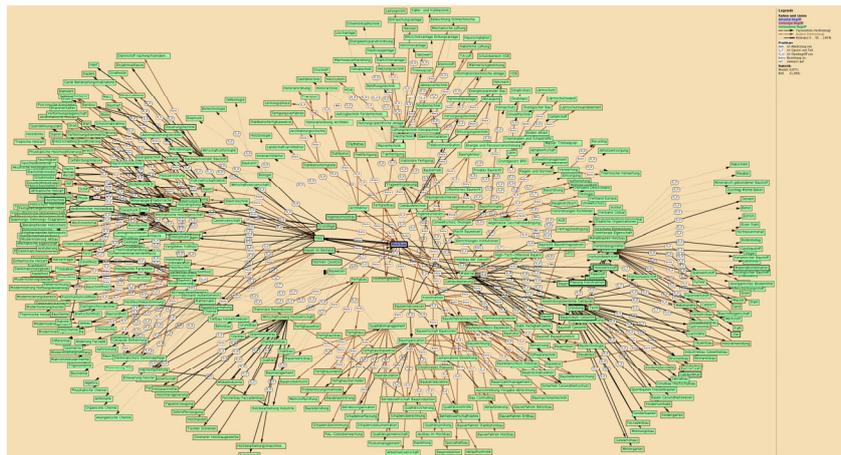


Abbildung 45: Beispiel für eine übermäßig umfangreiche Darstellung

In Bereichen starker Vernetzung wird die Darstellung jedoch schnell sehr umfangreich und unübersichtlich (Abbildung 45). Zudem wird die Antwortzeit unbefriedigend lange, da sich die Berechnung des Bildes, vor allem wegen der möglichst kollisionsfreien Anordnung der Begriffe und Prädikate, sehr aufwändig gestaltet.

Wie weiter oben bereits beschrieben, bietet das einzelne für die Suche vorgesehene Eingabefeld bereits sehr umfangreiche Formulierungsmöglichkeiten.

Darüber hinaus gibt es aber noch weitere einstellbare Optionen sowohl für das semantische Netz als auch für den Index (Abbildung 46).

Such-Optionen einstellen

Einstellungen

Begriffe-Netz	Suche
Darstellung	
<input type="radio"/> Darstellung aus <input type="radio"/> Graphische Darstellung <input type="radio"/> Listendarstellung <input checked="" type="radio"/> Darstellung graphisch und als Liste	
Minimale Relevanz	Anzahl Ergebnisse pro Seite <input type="text" value="10"/>
Maximale Entfernung	Kategorien
Taxonomie	<input type="checkbox"/> Lexikon
<input checked="" type="checkbox"/> Bereits aufgelistete Beziehungen unterdrücken	<input type="checkbox"/> Wiki
Sprachliche Alternativen <input type="text" value="41"/>	<input type="checkbox"/> Wissenschaftliche Suche
	<input type="checkbox"/> Patentrecherche
	<input type="checkbox"/> Produktsuche
	Themengebiete
	<input type="checkbox"/> Architektur
	<input type="checkbox"/> Holz
	<input type="checkbox"/> Holzbau
	<input type="checkbox"/> Rohstoff
	<input type="checkbox"/> Technologie
<input type="button" value="Einstellungen setzen"/>	

Abbildung 46: Dialog für die Such-Optionen

Für das semantische Netz existieren folgende Optionen:

- Nur graphische, nur textliche Listen-Darstellung, beides oder keine Darstellung
- Mit der minimalen Relevanz wird festgelegt, bis zu welcher Relevanz Begriffe und Beziehungen angezeigt werden sollen. Sobald die ausmultiplizierte Relevanz diesen Wert unterschreitet, werden das entsprechende und die folgenden Elemente nicht mehr in die Darstellung mit einbezogen
- Die maximale Entfernung bestimmt, wie weit die in Beziehung stehenden Begriff entfernt sein können, um in die Anzeige mit einbezogen zu werden
- Über die Taxonomie kann festgelegt werden, ob alle Elemente, nur Elemente der Systematik oder nur Elemente, die nicht zur Systematik gehören, mit einbezogen werden sollen
- Da die textliche Listen-Darstellung nur hierarchisch erfolgen kann, muss hierzu das Netz aufgelöst werden, was dazu führen kann, dass bestimmte Zweige mehrfach vorkommen. Durch diesen Schalter kann die wiederholte Auflistung von Zweigen unterbunden werden
- Die maximale Anzahl der aufgelisteten sprachlichen Alternativen

Außerdem existieren für den Index folgende Optionen:

- Anzahl der je Seite aufgelisteten Ergebnisse
- Beschränkung der Kategorien.
Falls nichts ausgewählt ist, besteht keine Beschränkung
- Beschränkung der Themengebiete.
Falls nichts ausgewählt ist, besteht keine Beschränkung

11.5 Redaktion

Ein Benutzer mit Redaktionsrechten kann je nach Höhe des vergebenen Rechtes das Lexikon und das semantische Netz bearbeiten:

- Stufe 0: keine Editierrechte
- Stufe 1: Editierrecht für Lexikon
- Stufe 2: zusätzliches Editierrecht für Begriffe und Beziehungen
- Stufe 3: zusätzliches Editierrecht für Prädikate

Auf die Redaktion des Lexikons, das auf JSPWiki basiert, wird nachfolgend nicht näher eingegangen³³.

11.5.1 Semantisches Netz

Bei vollem Editierrecht stellt sich die Redaktionsansicht wie in Abbildung 47 gezeigt dar. Um die Begriffe, Prädikate bzw. Beziehungen zu bearbeiten, klickt man auf die gewünschte Funktion und gelangt so in die entsprechende Auflistung.



IRIS Ein Informationssystem rund um den Holzbau

Angemeldeter Benutzer: **Max Mustermann**

Navigation

[Startseite](#)
[Anmelden](#) [Abmelden](#)

Informationen

- [Lexikon](#)
- [Datenbanken](#)

Redaktion

- [Semantisches Netz](#)

Aktuelles

Impressum

- [Weitere Informationen](#)
- [Glossar](#)
- [Rückmeldung](#)

Redaktion

Semantisches Netz

- [Begriffe im semantischen Netz anlegen und bearbeiten \(Aktuelle Anzahl: 14083\)](#)
- [Prädikate im semantischen Netz anlegen und bearbeiten \(Aktuelle Anzahl: 9\)](#)
- [Beziehungen im semantischen Netz anlegen und bearbeiten \(Aktuelle Anzahl: 4489\)](#)

Abbildung 47: Redaktionsfunktionen

³³ Eine Beschreibung mit Anleitung findet man im Internet unter: <http://www.jspwiki.org>

In Abbildung 48 ist exemplarisch die Auflistung der Begriffe³⁴ dargestellt. Oberhalb der eigentlichen Begriffe-Liste befindet sich der Filter, über den angegeben werden kann, welche Bedingungen die Einträge erfüllen müssen, um aufgelistet zu werden. Im dargestellten Beispiel müssen die Einträge im Namen die Zeichenfolge „holzbau“ enthalten.

Begriffe im semantischen Netz anlegen und bearbeiten Suche:

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Filter

Name:

gültig: Taxonomie: akzeptiert: löschbar: korrekt:

Limit Offset: 0 Limit Ergebnisanzahl: 100 Schrittweite: 50

Zurück [Vorwärts](#)

Begriffe

Es gibt 91 passende Einträge (von 14063 insgesamt). Gegenwärtig werden 91 Einträge gezeigt. [Daten exportieren](#).

Name (verwendet als)	glt. Tax. akz. lbr.	Relevanz				min max	var	Bearbeiten	Löschen
		Treffer seit	Kurzzeit seit	Langzeit seit					
Anschlussdetails Holzbauweise (Subjekt)	ja ja nein nein	0 30.01.2008	0 25.01.2008	0 04.01.2008	0 1	ja	<input type="button" value="Anlegen"/> <input type="button" value="Rücksetzen"/>	<input type="button" value="Eintrag"/>	
Bauteilanschluss Holzbau (Subjekt, Objekt)	ja ja nein nein	3 30.01.2008	0 25.01.2008	0 04.01.2008	0 1	ja	<input type="button" value="Eintrag"/>		
Bauverfahren Holzbau (Objekt)	ja nein ja nein	0 29.01.2008	0,018519 25.01.2008	0,5 24.01.2008	0 1	ja	<input type="button" value="Eintrag"/>		
Bauweise Holzbau (Objekt)	ja nein nein nein	1 28.01.2008	0,5 24.01.2008	0,5 23.01.2008	0,5 0,5	ja	<input type="button" value="Eintrag"/>		
Bauweise im Holzbau (Subjekt, Objekt)	ja ja nein nein	2 30.01.2008	0 25.01.2008	0,217391 04.01.2008	0 1	ja	<input type="button" value="Eintrag"/>		
Bemessung Holzbau (Subjekt, Objekt)	ja ja nein nein	47 30.01.2008	0 25.01.2008	0 04.01.2008	0 1	ja	<input type="button" value="Eintrag"/>		
Bemessungsgrundlage Holzbau	ja nein ja nein	3	0,5	0,5	0	ia			

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Abbildung 48: Auflistung der Begriffe des semantischen Netzes

Um einen Eintrag zu bearbeiten oder um seine Details einzusehen (Abbildung 49), klickt man in der Spalte „Bearbeiten“ auf den zugehörigen Link „Eintrag“.

Begriff bearbeiten

Name

gültig

Taxonomie

akzeptiert

löschbar

Relevanz

Treffer seit

Kurzzeit seit

Langzeit seit

minimal

maximal

variabel

Abbildung 49: Bearbeiten eines einzelnen Eintrags

Änderungen werden mit OK bestätigt oder mit Abbrechen verworfen. In beiden Fällen kommt man in die Listendarstellung zurück. Mit Rücksetzen werden nur die Feldinhalte auf die Ausgangswerte zurückgesetzt.

³⁴ Die Anzeige und Bearbeitung der Prädikate und Beziehungen funktioniert in derselben Weise und wird deshalb nicht separat ausgeführt.

11.6 Administration

Die Administration von IRIS umfasst:

- die Benutzerverwaltung
- die Partnerverwaltung
- die Verwaltung der Indexierung
- die Dokumentenverwaltung
- die Wortlistenverwaltung
- das Lernen
- die Verwaltung automatisierter Aufgaben

Um diese Aufgaben ausführen zu können, benötigt der Benutzer Administrationsrecht. Im Gegensatz zur Redaktion gibt es hier nur zwei Stufen:

- Stufe 0: Keine Administrationsrechte
- Stufe 1: Alle Administrationsrechte

Bei vollem Administrationsrecht stellt sich die Administrationsansicht wie in Abbildung 50 gezeigt dar. Um in eine bestimmte Funktion zu gelangen, klickt man auf den entsprechenden Link.

Navigation	Verwaltung
Startseite Anmelden Abmelden	
Informationen <ul style="list-style-type: none"> ▪ Lexikon ▪ Datenbanken 	Benutzer <ul style="list-style-type: none"> ▪ Anmelden ▪ Mein Passwort ändern ▪ Mein Benutzerkonto bearbeiten ▪ Benutzer auflisten und bearbeiten (Aktuelle Anzahl: 62) ▪ Neuen Benutzer anlegen ▪ Benutzerzugang freischalten
Redaktion <ul style="list-style-type: none"> ▪ Semantisches Netz 	Partner <ul style="list-style-type: none"> ▪ Neuen Projektpartner anlegen ▪ Projektpartner auflisten und bearbeiten (Aktuelle Anzahl: 5)
Verwaltung <ul style="list-style-type: none"> ▪ Benutzer ▪ Partner ▪ Indexierung und Suche ▪ Dokumente ▪ Wortlisten ▪ Lernen ▪ Automatisierte Aufgaben ▪ Verschiedenes 	Indexierung und Suche <ul style="list-style-type: none"> ▪ Index-Konfigurationen anlegen und bearbeiten (Aktuelle Anzahl: 26)
Aktuelles	Dokumente <ul style="list-style-type: none"> ▪ Dokumente (Aktuelle Anzahl: 13031)
Impressum <ul style="list-style-type: none"> ▪ Weitere Informationen ▪ Glossar ▪ Rückmeldung 	Wortlisten <ul style="list-style-type: none"> ▪ Wortlisten-Einträge gesamt (Aktuelle Anzahl: 0)
	Lernen <ul style="list-style-type: none"> ▪ Lernen im semantischen Netz
	Automatisierte Aufgaben <ul style="list-style-type: none"> ▪ Administration der automatisierten Aufgaben (Aktuelle Anzahl: 7)
	Verschiedenes <ul style="list-style-type: none"> ▪ Servlet-Container

Abbildung 50: Administrationsfunktionen

11.6.1 Zusammenhang zwischen Benutzern, Partnern und Index

Der Index in IRIS besteht normalerweise aus mehreren Teilen. Diese Teile werden über sogenannte Konfigurationen gesteuert. Jede dieser Konfigurationen kann wahlweise einem IRIS-Partner, der Inhalte zur Verfügung stellt, zugeordnet werden.

Wie weiter oben bereits gezeigt, kann man sich bei der Suche zu einem Element aus der Ergebnisliste Details anzeigen lassen. Zudem erhält man standardmäßig einen Link, um das zugehörige Dokument aufrufen zu können.

Falls die Quelle des Dokumentes über die Konfiguration mit einem Partner verbunden ist, dann kann der Aufruf des Dokumentes von einer Mitgliedschaft des angemeldeten Benutzers bei diesem Partner abhängen. Ein nicht angemeldeter Anwender kann dann über IRIS auf keinen Fall auf das entsprechende Dokument zugreifen. Eine Mitgliedschaft kann wahlweise zeitlich begrenzt oder unbegrenzt durch den Administrator eingetragen werden.

11.6.2 Benutzerverwaltung

Bis auf die Funktion „Benutzer auflisten und bearbeiten“ (Abbildung 51) stehen alle anderen Funktionen im Navigationsmenü über den Punkt „Anmelden“ für jeden Anwender zur Verfügung. Somit kann jeder Anwender sein Passwort und seine Kontaktdaten ändern. Der Administrator kann dies darüber hinaus für alle Benutzer durchführen. Jedoch besteht auch für den Administrator keine Möglichkeit, das aktuell gültige Passwort irgendeines Anwenders in Erfahrung zu bringen.

Benutzer auflisten und bearbeiten Suche:

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Filter

Benutzername: Vorname: Nachname: Emailadresse:

vollständige Datensätze
 unvollständige Datensätze
 vollständige und unvollständige Datensätze

Limit Offset: Limit Ergebnisanzahl: Schrittweite:

[Zurück](#) [Vorwärts](#)

Benutzer

Es gibt 62 passende Einträge (von 62 insgesamt). Gegenwärtig werden 62 Einträge gezeigt.

Benutzername Passwort	Vorname Nachname	Emailadresse Sprache	Stufen Admin, Edit, User	Gültig	Bearbeiten	Löschen
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	Nein	<input type="button" value="Anlegen"/>	<input type="button" value="Rücksetzen"/>
admin	Ford Perfekt	somewhere@space en	1, 3, 1	Ja	<input type="button" value="Eintrag
Passwort"/>	<input type="button" value="Löschen"/>
Al	Wolfgang Alversammer	alversammer@fh-rosenheim.de de	0, 0, 1	Ja	<input type="button" value="Eintrag
Passwort"/>	<input type="button" value="Löschen"/>
Andrea	Andrea Tausch	andrea.tausch@fobot.uni-freiburg.de de	0, 0, 1	Ja	<input type="button" value="Eintrag
Passwort"/>	<input type="button" value="Löschen"/>
asiegert	Albin Siegert	siegertalbin@gmx.de de	0, 0, 1	Ja	<input type="button" value="Eintrag
Passwort"/>	<input type="button" value="Löschen"/>
at	Andrea Till	andrea.till@gmx.de de	1, 3, 1	Nein	<input type="button" value="Eintrag
Passwort"/>	<input type="button" value="Löschen"/>
axer	Josef Axer	josef.axer@t-online.de de	0, 0, 1	Ja	<input type="button" value="Eintrag
Passwort"/>	<input type="button" value="Löschen"/>
Bauer	Markus ..	markus.moser@bauer-software.de ..	0, 0, 1	Ja	<input type="button" value="Eintrag
.."/>	<input type="button" value="Löschen"/>

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Abbildung 51: Benutzer auflisten und bearbeiten

Obwohl jeder Anwender sein Benutzerkonto bearbeiten kann (Abbildung 52), gibt es dennoch Einschränkungen bei den Rechten und den Mitgliedschaften, sofern der Anwender keine Administrationsrechte besitzt:

- Er kann seine Rechte nur verringern, jedoch nicht erhöhen
- Er kann keine Mitgliedschaften bei IRIS-Partnern eintragen

Benutzerkonto bearbeiten

Benutzername	mm
Anrede *	Herr
Titel	
Vorname *	Max
Nachname *	Mustermann
Adresszusatz	
Straße und Hausnummer *	Hochschulstraße 1
PLZ, Ort *	83024 Rosenheim
Land *	Deutschland
Telefonnummer	08031/805690
Weitere Telefonnummer	
Emailadresse *	mustermann@fh-rosenheim.d
Berufsgruppe	Forschung und Lehre
Sprache *	Deutsch
Verwaltungsstufe *	1: Alle Adminrechte
Redaktionsstufe *	3: Editierrecht für Prädikate
Gültig	<input checked="" type="checkbox"/>

Mitgliedschaften

Deutsche Gesellschaft für Holzforschung e.V. 14.09.2007 - 30.06.2008 Eintrag Löschen
--

[Hinzufügen](#)

Hinweise

- Wenn Sie nicht ausreichende Verwaltungsrechte besitzen, können Sie die Levels nur reduzieren.
- Verändern Sie die Rechte nur, wenn Ihnen die Auswirkung bewusst ist.
- Die Berechtigungen in der Auswahlliste nehmen absteigend zu. Ein Recht beinhaltet also auch das vorhergehende.
- Die mit einem * markierten Einträge sind Pflichtfelder.

Abbildung 52: Benutzerkonto bearbeiten

11.6.3 Partnerverwaltung

Projektpartner liefern Inhalte für IRIS. Bei der Indexierung können Dokumenten-Quellen mit Projektpartnern verbunden werden. Bei Benutzern können Mitgliedschaften bei Projektpartnern angegeben werden. Die bei Projektpartner hinterlegbaren Daten sind in Abbildung 53 ersichtlich. Die Verwaltung erfolgt über die Auflistung der Projektpartner (Abbildung 54).

Neuen Projektpartner anlegen

Einheit *	<input type="text" value="Bitte wählen..."/>
Firmenname *	<input type="text"/>
Firmenname 2	<input type="text"/>
Adresszusatz	<input type="text"/>
Straße und Hausnummer *	<input type="text"/>
PLZ, Ort *	<input type="text"/>
Land *	<input type="text" value="Bitte wählen..."/>
Telefonnummer	<input type="text"/>
Weitere Telefonnummer	<input type="text"/>
Telefax	<input type="text"/>
Emailadresse	<input type="text"/>
Homepage *	<input type="text"/>
Beschreibung	<div style="border: 1px solid #ccc; height: 40px; width: 100%;"></div>

Hinweis

- Die mit einem * markierten Einträge sind Pflichtfelder.

Abbildung 53: Neuen Projektpartner anlegen

Projektpartner auflisten und bearbeiten

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Filter ⓘ

Firmenname: Ort:

Limit Offset: Limit Ergebnisanzahl: Schrittweite:

Zurück [Vorwärts](#)

Partner

Es gibt 5 passende Einträge (von 5 insgesamt). Gegenwärtig werden 5 Einträge gezeigt.

Firmenname Firmenname 2	Adresse	Homepage	Optionen
<input type="text" value="Bitte wählen..."/>	<input type="text"/>	<input type="text"/>	<input type="button" value="Anlegen"/> <input type="button" value="Rücksetzen"/>
Deutsche Gesellschaft für Holzforschung e.V.	Bayerstr. 57 - 59 80335 München Deutschland	http://www.dgfh.de	<input type="button" value="Eintrag"/> <input type="button" value="Löschen"/>
Fachhochschule Rosenheim	Hochschulstrasse 1 83024 Rosenheim Deutschland	www.fh-rosenheim.de	<input type="button" value="Eintrag"/> <input type="button" value="Löschen"/>
HdZ	Hochschulstrasse 1 Hochschuls Rosenheim Deutschland	www.fh-rosenheim.de	<input type="button" value="Eintrag"/> <input type="button" value="Löschen"/>
Holzforschung Austria	Franz Grill-Str. 7 1030 Wien Österreich	www.holzforschung.at	<input type="button" value="Eintrag"/> <input type="button" value="Löschen"/>
ift Rosenheim	Theodor-Gietl-Strasse 55-57	www.ift-rosenheim.de	

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Abbildung 54: Projektpartner auflisten und bearbeiten

11.6.4 Indexierung

11.6.4.1 Konfigurationen

Der erste Schritt, um Dokumente in IRIS aufzunehmen, ist die Erstellung einer Konfiguration für diese Inhalte. Dazu ist in der Übersichtsmaske (Abbildung 55) der gesamte Pfad der gewünschten Konfigurationsdatei einschließlich Dateinamen anzugeben. Die geforderte Kurzbezeichnung für den Index (in der Maske „Index-Name“ genannt) dient bei der späteren Darstellung von Suchergebnissen als Quellenangabe zu einem Treffer. Optional kann der Indexkonfiguration ein Projektpartner zugeordnet werden. Alle mit Hilfe dieser Konfiguration dem System zugeführten Dokumenteneinträge werden dann als „diesem Projektpartner gehörend“ eingestuft. Damit ist auf Wunsch eine Beschränkung des Zugriffs auf diese Dokumente (oder auch nur ein Teil davon) auf jene Benutzer möglich, die für den betreffenden Projektpartner eine Mitgliedschaft eingetragen haben.

Beim Anlegen einer neuen Konfiguration ist es nicht erforderlich, dass die bezeichnete Datei bereits existiert.

Liste der Index-Konfigurationen Suche:

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Indexe

XML-Konfigurationsdatei	Prüfung	Index-Name	Projektpartner	Bearbeiten	Löschen
/opt/fisys/fisycrawlerconfig/	--		Kein Projektpartner	<input type="button" value="Anlegen"/> <input type="button" value="Rücksetzen"/>	
/opt/fisys/fisycrawlerconfig/CrawlingDGH.xml	<input checked="" type="checkbox"/> existiert <input checked="" type="checkbox"/> formal ok	DGH	Deutsche Gesellschaft für Holzforschung e.V.	<input type="button" value="Eintrag"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/WikiCrawling.xml	<input checked="" type="checkbox"/> existiert <input checked="" type="checkbox"/> formal ok	Wiki-System	(Kein Projektpartner)	<input type="button" value="Eintrag"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/BFAHamburg.xml	<input checked="" type="checkbox"/> existiert <input checked="" type="checkbox"/> formal ok	BFA	(Kein Projektpartner)	<input type="button" value="Eintrag"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/HAF.xml	<input checked="" type="checkbox"/> existiert <input checked="" type="checkbox"/> formal ok	Holzabsatzfonds	(Kein Projektpartner)	<input type="button" value="Eintrag"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/EuroLignum.xml	<input checked="" type="checkbox"/> existiert <input checked="" type="checkbox"/> formal ok	EuroLignum	(Kein Projektpartner)	<input type="button" value="Eintrag"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/BauenMitHolz.xml	<input checked="" type="checkbox"/> fehlt <input checked="" type="checkbox"/> nicht ok	bauenmitholz.de	(Kein Projektpartner)	<input type="button" value="Datei erzeugen"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/HolzHandel.xml	<input checked="" type="checkbox"/> fehlt <input checked="" type="checkbox"/> nicht ok	holzhandel.de	(Kein Projektpartner)	<input type="button" value="Datei erzeugen"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/CrawlingInfodienstHolz.xml	<input checked="" type="checkbox"/> existiert <input checked="" type="checkbox"/> formal ok	Informationsdienst-Holz	(Kein Projektpartner)	<input type="button" value="Eintrag"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/Holz.xml	<input checked="" type="checkbox"/> fehlt <input checked="" type="checkbox"/> nicht ok	holz.de	(Kein Projektpartner)	<input type="button" value="Datei erzeugen"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/ViFaHolz.xml	<input checked="" type="checkbox"/> fehlt <input checked="" type="checkbox"/> nicht ok	ViFaHolz	(Kein Projektpartner)	<input type="button" value="Datei erzeugen"/>	<input type="button" value="Löschen"/>
/opt/fisys/fisycrawlerconfig/DBIS.xml	<input checked="" type="checkbox"/> fehlt	DBIS	(Kein Projektpartner)	<input type="button" value="Datei erzeugen"/>	<input type="button" value="Löschen"/>

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).
[Zurück](#)

Abbildung 55: Übersicht der Indexierungs-Konfigurationen

Für alle bekannten Einträge wird eine doppelte Plausibilitätsprüfung durchgeführt und das Ergebnis dargestellt: ob die bezeichnete Konfigurationsdatei existiert und ob diese formal korrekt ist. Für den Fall, dass die Datei noch nicht existiert, bietet das System dem Nutzer an, eine neue Datei zu erzeugen. Ansonsten kann die Datei mit Hilfe des Links „Eintrag“ direkt angesehen und bearbeitet werden. Auch eine zuvor neu erzeugte, noch leere Datei kann auf diese Weise bearbeitet werden (vgl. Abbildung 56).

Index-Konfigurationen editieren Suche:

[Zurück zur Liste](#)

Konfigurationsdatei-Prototypen: [Muster-Konfiguration für das Erfassen von Dokumenten im System](#)
[Muster-Konfiguration NUR für das interne Crawling der Datenbank](#)
[Muster-Konfiguration NUR für das Wiki-Crawling](#)

Details	
XML-Konfigurationsdatei	/opt/isis/iisycrawlerconfig/EuroLignum.xml
Index-Name	EuroLignum
Projektpartner	Kein Projektpartner
Inhalt der Konfigurationsdatei:	
<input type="button" value="OK"/> <input type="button" value="Abbrechen"/> <input type="button" value="Rücksetzen"/>	
Fehlermeldung vom Auslesen der XML-Datei	Parsing XML failed: /opt/isis/iisycrawlerconfig/EuroLignum.xml

[Zurück zur Liste](#)

Abbildung 56: Editieren einer noch leeren Indexierungs-Konfigurationsdatei

Um dem Anwender das Erstellen einer neuen Konfiguration zu erleichtern, gibt es drei Vorlagen als Muster für verschiedene Zwecke:

- Erfassen neuer Dokumente
 Der Standardfall für die Aufnahme neuer Inhalte in IRIS. Die Vorlage ist so aufgebaut, dass damit Schritt 1 des IRIS-Crawling-Prozesses umgesetzt wird, d. h. insbesondere, dass zu den gefundenen Dokumenten Einträge in die Datenbank geschrieben werden, statt direkt in einen Lucene-Index. Der Administrator bzw. Redakteur muss die Vorlage noch gemäß der konkret zu erfassenden Datenquellen anpassen.
- Internes Crawling
 Im System muss es genau einen Eintrag mit dieser Konfiguration geben. Er realisiert Schritt 2 des IRIS-Crawling-Prozesses. Die Vorlage ist so konzipiert, dass sie möglichst ohne Änderungen direkt übernommen werden kann.
- Wiki-Crawling
 Mit dieser Vorlage kann mit Hilfe der Erweiterungen zu Regain das JSPWiki-System von IRIS durchlaufen werden. Die Vorlage sollte ohne Anpassungen direkt übernommen werden können. Wie für das „Interne Crawling“ genügt ein einziger Eintrag mit dieser Konfiguration.

Die Vorlagen öffnen sich jeweils in einem Popup-Fenster (Abbildung 57) und können somit leicht vom Bearbeiter in das noch leere Textfeld umkopiert werden.

```

<?xml version="1.0" encoding="ISO-8859-1"?>

<!--
| Vorgehensweise zur Erstellung einer Konfiguration:
| 1. Vergewissern Sie sich, dass vor dem einleitenden <?xml ...> tag ke
| 2. Gehen Sie anhand der durchnummerierten Punkte durch die Datei und
| Ihre Einstellungen vor. Schauen Sie auf die Beschreibungen in den
| zu jedem dieser Punkte.
| 3. Vergewissern Sie sich, dass der <configuration> tag vollständig is
| schließende Ende-tag </configuration> muss am Ende der Datei vorha
| die anderen Tags umschließen)
|
| =====
| How to create a proper configuration:
| 1. Make sure there is no text before the introducing <?xml ...> tag
| 2. Go through the file and make your settings along the serially numb
| Have a look at the description comments at every of these.
| 3. Make sure the <configuration> is complete (the end tag </configura
| at the end of the file and surrounds the other tags)
|
| --->

<!--
| 0. DOCTYPE Angaben. Einfach stehen lassen wie es ist.
| =====
| 0. DOCTYPE definition. Just leave it untouched.
|
| --->
<!DOCTYPE configuration [
<!ENTITY amp "&";>
<!ENTITY lt "<";>
<!ENTITY minus "E";>
]>

<!--
| Konfiguration des regain crawler (um einen Suchindex zu erstellen)
|
| You can find a detailed description of all configuration tags here:
| http://regain.murfmman.de/wiki/de/index.php/CrawlerConfiguration.xml
| (Allerdings wird man nur zur englischen Beschreibung verwiesen)
|
| Das hier ist eine modifizierte Version mit erweiterten IISY Optionen!
| Achten Sie daher besonders auf die Kommentare der mit ERWEITERUNG mar
| =====
| Configuration for the regain crawler (for creating a search index)
|
| You can find a detailed description of all configuration tags here:
| http://regain.murfmman.de/wiki/en/index.php/CrawlerConfiguration.xml

```

Abbildung 57: Ausschnitt aus einer Indexierungs-Konfigurations-Vorlage

Die Vorlagen sind als HTML realisiert und Ergebnis einer Transformation aus den zugrunde gelegten XML-Dateivorlagen. Dadurch sind in den Vorlagen die reichlich beigefügten Kommentare farblich abgesetzt. Es ist zu beachten, dass die Farbgebung beim Einfügen in das für die Bearbeitung vorgesehene Textfeld jedoch verloren geht (vgl. hierzu auch Abbildung 58).

Der innere Aufbau der Vorlagen ist so ausgelegt, dass der Bearbeiter die Konfiguration vom Anfang bis zum Ende Punkt für Punkt durchgehen sollte. An verschiedenen Stellen sind dann gemäß den jeweiligen Bedürfnissen die benötigten Anpassungen vorzunehmen. Die grundsätzliche Struktur entstammt der in „Regain“ enthaltenen Musterkonfiguration. Sie wurde jedoch durch zahlreiche Kommentare zur Verwendung angereichert. Insbesondere die über „Regain“ hinausreichenden Möglichkeiten werden ausführlich als „Inline-Dokumentation“ beschrieben.

IRIS benötigt eine Verbindung zu seiner Datenbank. In den Vorlagen sind die Verbindungen stets für die während der Entwicklung eingesetzte MySQL-Datenbank konfiguriert. Falls bei der Installation von IRIS abweichend eine andere Datenbank verwendet wird oder angesprochen werden muss, sind die in den Vorlagen enthaltenen Konfigurationen zur Datenbank-Verbindung an die konkret vorhandene Systemumgebung anzupassen.

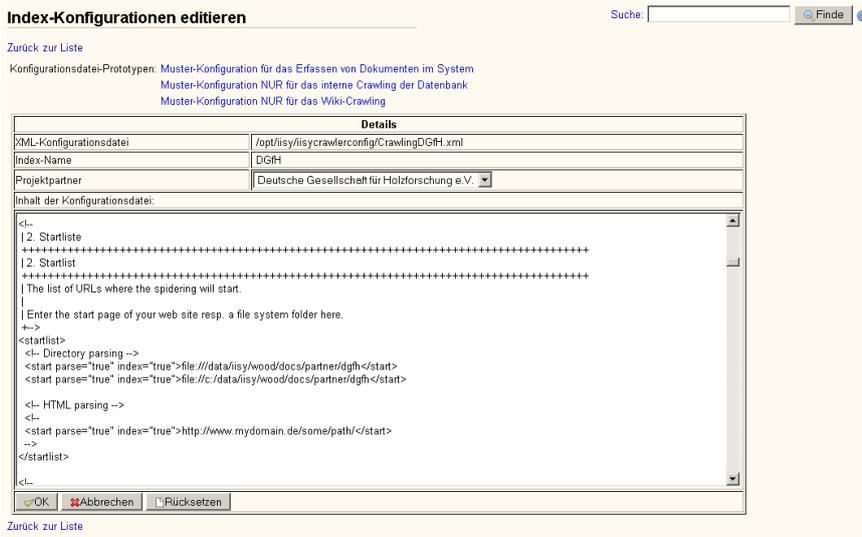


Abbildung 58: Editieren einer vorhandenen Indexierungs-Konfigurationsdatei

Unabhängig davon, ob die Konfiguration mit Hilfe einer der Vorlagen erstellt wurde oder nicht, kann sie mit der Editiermaske bearbeitet werden. In dieser Maske können der Textinhalt der Konfigurationsdatei sowie der zugeordnete Projektpartner geändert werden (Abbildung 58).

Beim Aufruf der Maske wird der Dateiinhalt der angegebenen Datei eingelesen und im Textfeld dargestellt. Beim Einlesen wird zudem eine Prüfung auf formale Korrektheit durchgeführt. Sollte das System hier einen Fehler feststellen, gibt es als Hilfestellung für den Bearbeiter eine kleine Fehlermeldung am Bildschirm unterhalb des Textfeldes aus (vgl. z. B. Abbildung 56 unten: die angegebene Datei ist dort noch leer und erfüllt damit zum Zeitpunkt des Aufrufes der Maske nicht die formalen Anforderungen an eine syntaktisch korrekte Konfiguration). Mit dem „OK“-Knopf kann der neue Text nach der Bearbeitung abgespeichert werden. Hierbei wird im Anschluss daran die Maske neu geladen.

11.6.4.2 Starten des Crawling-Prozesses

Das Crawling wird von einem separaten Prozess erledigt, der unabhängig von der Webanwendung IRIS als eigenständiges Programm läuft. Der „IRIS-Crawler“ kann bei der Übersetzung der Quelldateien als Jar-Datei (Java-Bibliothek) samt aller weiteren benötigten Dateien (Skripten, Präparatoren-Plugins, Logging-Konfiguration) in einem eigenen Verzeichnis erzeugt und zusammengestellt werden (vgl. hierzu auch Anhang 15.5).

11.6.4.2.1 Aufruf von der Kommandozeile

Bei Bedarf kann der Crawling-Prozess von der Kommandozeile aus per Hand gestartet werden. Der Aufruf erfolgt als Java-Programm mit den üblichen Parametern zu Java selbst und den Programmparametern von „Regain“. An Letzteren wurden keinerlei Modifikationen vorgenommen, so dass für eine Erläuterung auf das Regain Anwenderhandbuch ([Schn3]) verwiesen wird.



Abbildung 59: Aufruf des Crawlers von der Kommandozeile

In Abbildung 59 wird beispielhaft ein möglicher Aufruf des Crawlers unter Windows dargestellt. Im Folgenden sollen die einzelnen Bestandteile des gezeigten Kommandos kurz erläutert werden:

Das Kommando beginnt mit dem Start der Java Virtual Machine (JVM) durch den Aufruf `java`. Ihm folgen optionale Parameter für die JVM, hier neue Speichereinstellungen (`-Xms64m -Xmx640m -Xss512k`). Sie sorgen dafür, dass dem Crawler mehr Hauptspeicher zur Verfügung steht, als nach den Voreinstellungen. Allerdings sind die in diesem Beispiel genannten Optionen nicht standardisiert. Es folgt die Benennung der Java-Bibliothek, die zur Ausführung kommen soll (`-jar regain-crawler.jar`). Damit der Crawler ordnungsgemäß ablaufen kann, muss ihm mindestens noch die zu verwendende Konfigurationsdatei mitgeteilt werden (`-config /opt/iisy/iisycrawlerconfig/IRB.xml`). Mit Hilfe weiterer optionaler Parameter kann das Verhalten noch genauer gesteuert werden (hier z. B. `-retryFailedDocs`). Für die möglichen Zusatzparameter und eine ausführliche Erläuterung Derselben wird auf das Regain Anwenderhandbuch ([Schn3]) verwiesen.

Es sei darauf hingewiesen, dass der manuelle Aufruf über die Kommandozeile in IRIS nicht als der übliche Weg zum Starten des Crawling-Prozesses vorgesehen ist. Vielmehr besitzt IRIS ein Modul samt Weboberfläche zur automatisierten, zeitgesteuerten Planung von Crawling-Durchläufen (vgl. Kapitel 11.6.4.2.2). Dieses sollte nach Möglichkeit benutzt werden.

11.6.4.2.2 Zeitgesteuerter Aufruf über Quartz

Quartz ist ein Java Framework zur Terminierung von Aufgaben. Im Umfeld von IRIS wird es genutzt, um eine Zeitsteuerung von Prozessen zu realisieren, vor allem für Durchführung von regelmäßig wiederkehrenden Crawling-Vorgängen. Zum Einsatz kommt Quartz derzeit in der Version 1.5.2.

Quartz besteht aus zwei logischen Bestandteilen: einem Scheduler Server und einem oder mehreren Clients. Der Server verwaltet die Liste der ihm gestellten Aufgaben („Jobs“ genannt), die zugehörigen Zeiteinstellungen („Trigger“), d. h. wann welche Aufgabe zur Ausführung kommen soll, und übernimmt den tatsächlichen Start dieser Arbeiten. Der Client (oder die Clients) sind dafür zuständig, dem Server neue Aufgaben und/oder neue Zeitvorgaben zu übertragen, gegebenenfalls angefallene Ergebnisse abzuholen und nicht mehr benötigte Aufgaben von der Liste entfernen zu lassen.

Der Quartz-Server läuft unabhängig von IRIS als separater Prozess in einer eigenen Instanz der JVM. Die IRIS-Anwendung spricht ihn als Client an. Die Kommunikation zwischen den beiden geschieht mittels RMI (Remote Method Invocation). Auf eine Beschreibung dieser Technik wird in diesem Dokument verzichtet und stattdessen auf zahlreiche Quellen im Internet hierzu verwiesen.

Bei IRIS besteht ein Job darin, den Crawling-Vorgang zeitgesteuert regelmäßig anzustoßen. Es geht also um die Verknüpfung eines Kommandos zum Start des IRIS-Crawlers (vgl. Kapitel 11.6.4.2.1) mit einer Zeitkomponente („Job Scheduling“). Der Quartz-Server bringt das Kommando dann zur Ausführung.

Der „IRIS Scheduler Server“ kann wie der „IRIS-Crawler“ als Jar-Datei samt aller nötigen weiteren Bibliotheken und Dateien in einem eigenen Verzeichnis bei der Übersetzung zusammengestellt werden. Damit die nachfolgenden Ausführungen funktionieren, muss er als Hintergrundprozess laufen und für die IRIS-Anwendung ansprechbar sein. Beim Aufruf der entsprechenden Mas-

ken wird dies geprüft. Bei Erfolglosigkeit erhält der Benutzer einen Hinweis wie in der folgenden Abbildung 60 dargestellt.

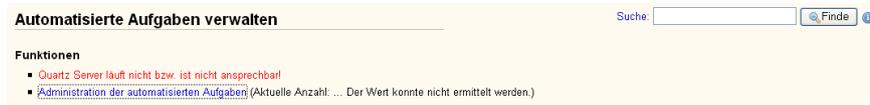


Abbildung 60: Quartz Scheduler Server nicht ansprechbar

Der Scheduler Server kann unter Windows von dem Verzeichnis aus, in dem sich die Dateien befinden, mit dem Kommando `start javaw -jar scheduler.jar` erfolgen. Für Linux ist bei IRIS das Shell Script `iisysched` beigefügt. Mit den üblichen Parametern `start`, `stop` oder `status` kann damit der Scheduler Server gesteuert und überwacht werden. Der Scheduler Server kann auch gestoppt werden, indem man die Datei `/tmp/schedserver.run` löscht, die er gewissermaßen als Bestätigung beim Start regelmäßig erstellt.

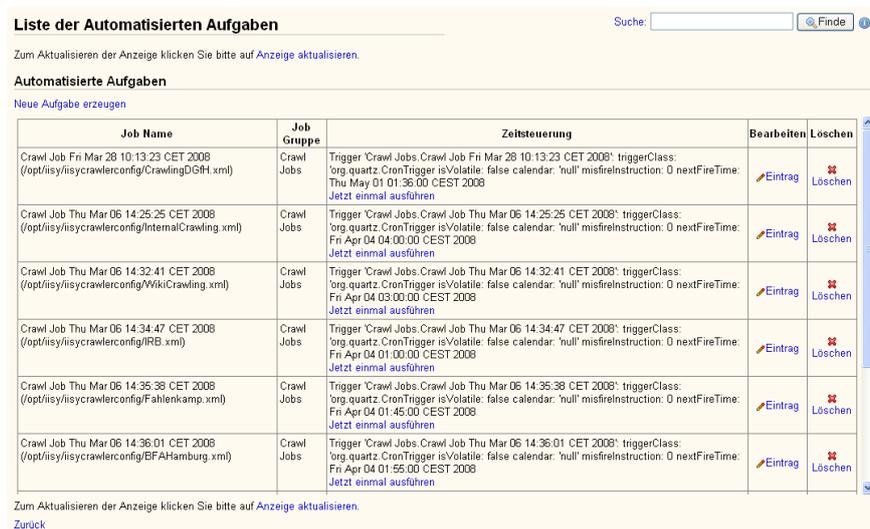


Abbildung 61: Übersicht der Automatisierten Aufgaben

In Abbildung 61 ist beispielhaft eine Auflistung von konfigurierten Aufgaben dargestellt. Von Quartz her müssen die einzelnen Aufgaben (Jobs) namentlich benannt und in Gruppen eingeteilt werden. Als Name wird in IRIS standardmäßig „Crawl Job“ gefolgt vom Zeitpunkt der Eintragung vergeben. Die Angabe in Klammern gibt als Erleichterung für den Benutzer die Bezeichnung der Konfigurationsdatei, auf die sich der Crawl-Job bezieht, wieder, ist aber nicht Teil des eigentlichen Namens. Die Aufgaben werden der Gruppe „Crawl Jobs“ zugeteilt. In der Spalte Zeitsteuerung werden die Daten wiedergegeben, die das Quartz-Zeitsteuerungsobjekt (Trigger) von sich preisgibt. Hier kann insbesondere die Angabe „nextFireTime“ interessant sein, die den Zeitpunkt benennt, an dem der Job das nächste Mal ausgeführt werden wird (in englischer Notation). In derselben Spalte gibt es zudem einen Link, mit dem der Vorgang einmalig außerplanmäßig sofort gestartet werden kann.

Das Anlegen und die Konfiguration einer neuen Aufgabe geschieht über den Link „Neue Aufgabe erzeugen“. In der sich daraufhin öffnenden Maske (siehe Abbildung 62) sind alle relevanten Daten einzutragen.

Abbildung 62: Neue Aufgabe festlegen

Der Eintrag unter „Java Kommando“ kann theoretisch jedes ausführbare Kommando enthalten. Standardmäßig wird dem Nutzer der in Abbildung 62 gezeigte Vorschlag gemacht, der auf ein in IRIS enthaltenes Script (Linux) bzw. eine Batch-Datei (Windows) verweist. Das Script bzw. die Batch-Datei enthält alle notwendigen Bestandteile zum Aufruf des Crawlers von der Kommandozeile (vgl. Kapitel 11.6.4.2.1) außer den Namen der Konfigurationsdatei und möglicher weiterer Optionen, die als Parameter zu übergeben sind. Unter Linux ist zu beachten, dass die genannte Script-Datei ausführbar sein muss, d. h. das Dateiattribut **execute** besitzen muss.

Bei „Konfigurationsdatei“ ist die bei dieser Aufgabe zu benutzende Crawler-Konfigurationsdatei zu benennen. Standardmäßig werden dem Anwender die in IRIS eingetragenen Konfigurationen (vgl. Kapitel 11.6.4.1) zur Auswahl angeboten. Zur Vereinfachung können in der Maske die optionalen Regain-Parameter zur genaueren Steuerung des Crawling-Vorganges im Formularblock darunter direkt per Häkchen ausgewählt werden (vgl. Kapitel 11.6.4.2.1).

Im Block „Zeitsteuerung“ wird festgelegt, wann das Kommando und damit der Crawling-Lauf jeweils auszuführen ist. Die Syntax des zum Einsatz kommenden so genannten CronTrigger von Quartz orientiert sich eng an dem weit verbreiteten und zu ähnlichen Zwecken genutzten Linux-Programm **cron**. Im Wesentlichen wird ein Zeitpunkt festgelegt, bei dem bestimmte Teile per Platzhalter (Stern bzw. Fragezeichen) frei bleiben. Beispielsweise würde die Festlegung „jeden Tag um 2:15 Uhr“ wie folgt aussehen: 0-15-2-?-*-*-. Auf eine ausführliche Darstellung der Syntax wird an dieser Stelle jedoch verzichtet und stattdessen auf die Erklärung des CronTrigger im Internet unter <http://quartz.sourceforge.net/javadoc/org/quartz/CronTrigger.html> oder auch <http://wiki.opensymphony.com/display/QRTZ1/CronTriggers+Tutorial> verwiesen. Beide Quellen sind in englischer Sprache gehalten. Die letztgenannte ist IRIS auch als Hilfetext beigefügt. Hilfreich zum Verständnis könnte auch das Studium der Funktionsweise des weiter oben genannten Programmes **cron** sein.

Beim Drücken des „OK“-Knopfs wird die neue Aufgabe mit den genannten Daten an den Scheduler Server übermittelt, um sie dort zu registrieren.

Automatisierte Aufgabe bearbeiten Suche: holz

[Zurück zur Liste](#)

Details															
Java Kommando	<input type="text" value="/opt/isisysycrawler/bin/crawl"/>														
Java Parameter (optional)	<input type="text"/>														
Konfigurationsdatei	<input type="text" value="/opt/isisysycrawlerconfig/InternalCrawling.xml"/>														
Regain Parameter (Indexierungstool, optional)	<input type="checkbox"/> -forceNewIndex <input checked="" type="checkbox"/> -retryFailedDocs <input type="checkbox"/> -onlyEntries <input type="text"/> <input type="checkbox"/> -logConfig <input type="text"/>														
Zeitsteuerung	Hilfe zur Festlegung der Zeitsteuerung (engl.) (Quelle: http://wiki.opensymphony.com/display/ORTZ1/CronTriggers+Tutorial) <table border="1"> <thead> <tr> <th>Sekunden</th> <th>Minuten</th> <th>Stunden</th> <th>Tag im Monat</th> <th>Monat</th> <th>Tag der Woche</th> <th>Jahr</th> </tr> </thead> <tbody> <tr> <td><input type="text" value="0"/></td> <td><input type="text" value="0"/></td> <td><input type="text" value="4"/></td> <td><input <="" td="" type="text" value="?"/> <td><input type="text" value="^"/></td> <td><input type="text" value="^"/></td> <td><input type="text"/></td> </td></tr> </tbody> </table>	Sekunden	Minuten	Stunden	Tag im Monat	Monat	Tag der Woche	Jahr	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="4"/>	<input <="" td="" type="text" value="?"/> <td><input type="text" value="^"/></td> <td><input type="text" value="^"/></td> <td><input type="text"/></td>	<input type="text" value="^"/>	<input type="text" value="^"/>	<input type="text"/>
Sekunden	Minuten	Stunden	Tag im Monat	Monat	Tag der Woche	Jahr									
<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="4"/>	<input <="" td="" type="text" value="?"/> <td><input type="text" value="^"/></td> <td><input type="text" value="^"/></td> <td><input type="text"/></td>	<input type="text" value="^"/>	<input type="text" value="^"/>	<input type="text"/>									

[Zurück zur Liste](#)

Abbildung 63: Editieren einer vorhandenen Aufgabe

Das Bearbeiten einer bereits vorhandenen Aufgabe geschieht im Prinzip auf die gleiche Weise wie beim Neuanlegen. Der einzige wesentliche Unterschied ist, dass die zuletzt eingetragenen Werte beim Aufruf der Maske bereits gesetzt sind (siehe Abbildung 63).

Nachdem man die gewünschten Änderungen im Formular vorgenommen hat, wird mit dem Drücken des „OK“-Knopfes der Änderungsvorgang ausgelöst. Da es von Quartz her keine Möglichkeit gibt, dies sauber direkt in einem Zug zu erledigen, sind zwei Schritte notwendig, die unmittelbar nacheinander ausgeführt werden: die vorhandene Aufgabe wird von der Aufgabenliste des Scheduler Server entfernt und anschließend mit den in der Maske eingetragenen Werten eine neue Aufgabe erzeugt und eingetragen.

11.6.5 Dokumentenverwaltung

Hierbei handelt es sich um die Verwaltung von Dokumenteneinträgen, das heißt um die Angaben, die dem System über (!) Dokumente vorliegen. Die einzelnen Einträge werden im Zuge des IRIS-Crawling Schritt 1 (vgl. Kapitel 10.4.5) in die Datenbank geschrieben. Die Verwaltungsfunktion dient vor allem zur Kontrolle sowie für manuelle Korrekturen bei einzelnen Einträgen.

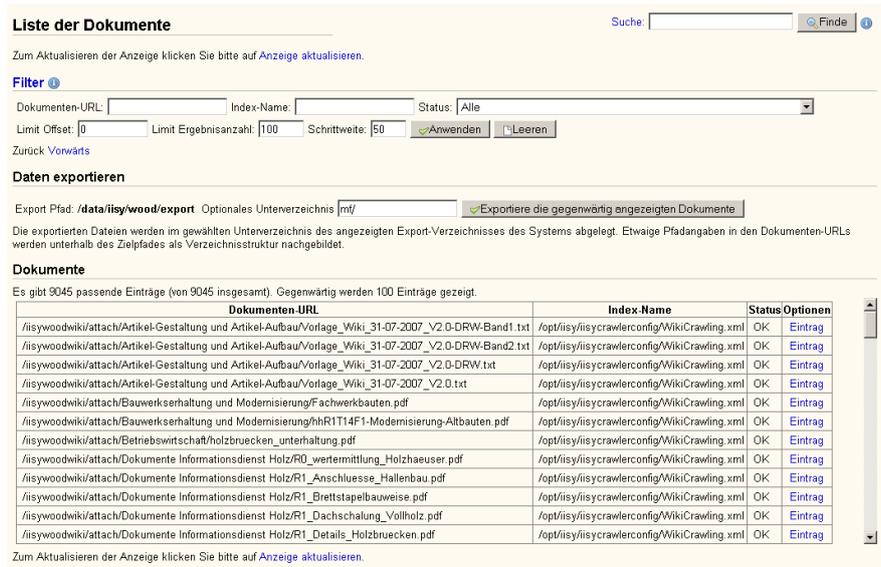


Abbildung 64: Überblick der Dokumenten-Einträge

Zunächst gelangt man auf die Übersichtsmaske der Dokumente (siehe Abbildung 64). Hier können im Zusammenspiel mit den Filterungsmöglichkeiten der Anzeige grundlegende Angaben zu den in der Datenbank vorhandenen Einträgen eingesehen werden. Insbesondere lassen sich auf einfache Weise die mit Hilfe einer bestimmten Index-Konfiguration gefundenen Dokumente sowie deren Status einsehen.

Mit Hilfe der Exportfunktion können die Inhalte der in der Maske ausgewählten Dokumenten-Einträge als Textdateien in eine Verzeichnisstruktur unterhalb des Exportverzeichnisses geschrieben werden. Dies kann für externe Auswertungszwecke interessant sein.

Unter „Eintrag“ lassen sich die Details zum entsprechenden Dokument einsehen und teilweise bearbeiten (siehe Abbildung 65).

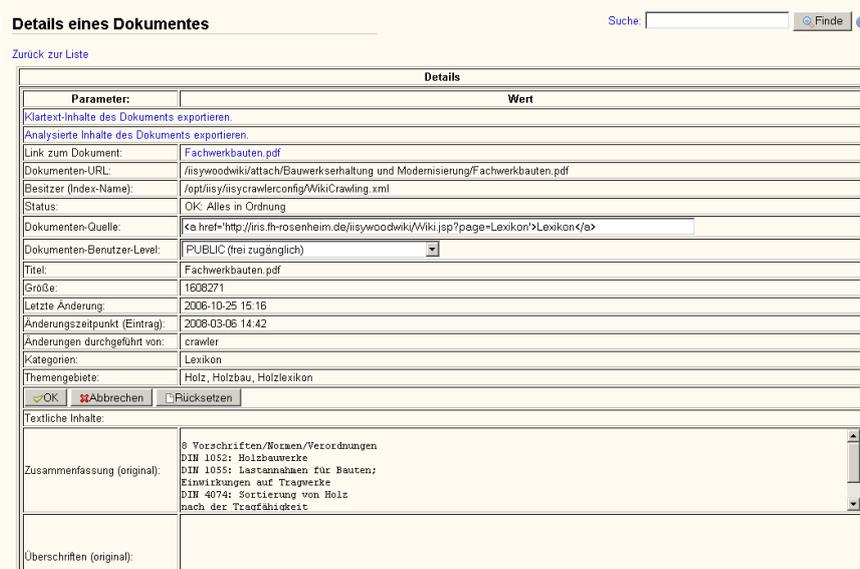


Abbildung 65: Details eines Dokumenten-Eintrages

Die wesentlichen Attribute (neben weiteren für technische Zwecke), die mit jedem Eintrag in der Datenbank gespeichert werden, sind in folgender Übersichtstabelle aufgelistet:

Tabelle 6: Attribute eines Dokumenten-Eintrages

Attribut	Automatisch	Veränderlich	Manuell änderbar
URL	Ja	Nein	Nein
Index-Name	Ja	Nein	Nein
Status	Ja	Ja	Nein
Letzte Modifikation (Datei)	Ja	Ja	Nein
Letzte Modifikation (Eintrag)	Ja	Ja	Nein
Letzte Modifikation durch	Ja	Ja	Nein
Größe	Ja	Ja	Nein
Quelle	Nein *	Ja	Ja
Benutzer-Level (Zugriffsschutz)	Ja *	Ja	Ja
Zugeordnete Kategorien	Nein *	Ja	Ja
Zugeordnete Themengebiete	Nein *	Ja	Ja
Titel	Ja	Ja	Nein
Zusammenfassung	Ja	Ja	Nein
Überschriften	Ja	Ja	Nein
Inhalt (Fließtext)	Ja	Ja	Nein

*) Mit Hilfe einer über den Standard von „Regain“ hinausgehenden Erweiterung in der Indexkonfiguration können optional automatisch zu vergebende Werte für Einträge vergeben werden (siehe dazu Kapitel 11.7). Für den Benutzer-Level wird im Falle von Web-Dokumenten standardmäßig der Level 0, sonst Level 2 vergeben (zur Bedeutung siehe weiter unten).

Dabei bedeutet **Automatisch**, dass der Wert standardmäßig vom System gesetzt wird. Die Spalte **Veränderlich** zeigt auf, welche Attribute sich im Laufe der Zeit verändern können. **Manuell änderbar** sind solche Daten, die der Administrator bzw. Redakteur unmittelbar manipulieren kann. Die veränderlichen Attribute sind auch ein wesentlicher Grund für den Zwischenschritt bei der Indexierung in IRIS und die Existenz dieser Dokumenten-Einträge überhaupt (vgl. hierzu auch die Ausführungen in Kapitel 10.4.5).

Die Attribute **Titel**, **Zusammenfassung**, **Überschriften** und **Inhalt** werden jeweils in zwei Fassungen abgelegt: einmal als original und einmal in analysierter Form. „Original“ sind dabei die erkannten reinen Texte ohne jegliche Grafiken, Formatierungen usw. Das ursprüngliche Dokument lässt sich daraus im Allgemeinen nicht rekonstruieren. Die analysierte Form („stemmed“) sind die aus den vorgenannten „Original“-Texten ermittelten Stammformen, die mit Hilfe eines Analysierers gewonnen wurden. Die Angaben dienen vor allem der Qualitätskontrolle der in IRIS eingestellten Inhalte. Damit lässt sich beispielsweise überprüfen, ob die Dokumente aus einer bestimmten Quelle korrekt ausgelesen werden konnten oder ob möglicherweise nur „Datenmüll“ extrahiert wird. Auch können diese Angaben wertvolle Hinweise zur Überprüfung der Plausibilität des Ergebnisses einer Suchanfrage liefern.

Eine für den Administrator bzw. Redakteur sicherlich herausragend wichtige Einstellung ist der Dokumenten-Benutzer-Level. Dieser gibt an, wer Zugriff auf das zugrunde liegende Dokument bekommt. Derzeit gibt es drei mögliche Stufen:

Tabelle 7: Berechtigungsstufen für Dateizugriff (Benutzer-Level)

Benutzer-Level	Zugriff
0 (PUBLIC)	Alle Benutzer
1 (RESTRICTED)	Alle bei IRIS angemeldeten Benutzer
2 (PRIVATE)	Nur solche Benutzer, die als gültiges Mitglied beim dem Partner eingetragen sind, der der Index-Konfiguration des Dokumentes zugeordnet ist

Die einem Dokumenten-Eintrag zugeordnete Index-Konfiguration ist diejenige, mit deren Hilfe der Eintrag beim IRIS-Crawling in die Datenbank aufgenommen worden ist. Sie ist im Attribut **Index-Name** festgehalten. Man kann es so interpretieren, dass das Dokument „zu dieser Konfiguration gehört“.

Damit das Attribut **Quelle** nicht missverstanden wird, sei es hier kurz erläutert. Damit ist nämlich nicht die Original-Adresse des Dokumentes gemeint (diese Angabe steht schon im Attribut URL). Vielmehr ist hier die eigentliche Bezugsquelle eines Dokumentes zu hinterlegen. Das kann beispielsweise die Homepage des Eigentümers eines Dokumentes sein. Vor allem kann das aber bei Dokumenten, die nicht einfach direkt vom Server oder aus dem Internet heruntergeladen werden können bzw. dürfen, die Adresse sein, unter der das Original-Dokument käuflich erworben werden kann.

11.6.6 Wortlistenverwaltung

Im System werden Wortlisten geführt, um die Indexierung im Laufe der Zeit zu optimieren. Sie können während der Indexierung im Zuge des Einlesens des Wortstromes, d. h. der rein textlichen Inhalte eines Dokuments, mit Hilfe von speziell für IRIS modifizierten Analysierern geschrieben werden. Es werden dabei alle neu gefundenen Wörter in die Liste aufgenommen. In Abhängigkeit von der lexikalischen Struktur des Wortes wird ihm einer von mehreren möglichen Status automatisch zugewiesen. Jedes Wort kann nur genau einen Status haben. Obwohl technisch gesehen nur eine einzige Liste existiert, kann man sie gedanklich auch trennen. Alle Wörter mit gleichem Status bilden dann von der Vorstellung her jeweils eine eigene Wortliste.

Die Wortlisten können als Grundlage für weitergehende Auswertungen, als Übersicht der tatsächlich in erfassten Texten vorkommenden Begriffe sowie zur Überprüfung der Wortstammermittlung und der Verbesserung von Analysierern benutzt werden. Eine Anwendung könnte zudem darin bestehen, aus den Wortlisten Kandidaten für das semantische Netz zu extrahieren.

Einen Überblick über die verschiedenen Status gibt die folgende Tabelle.

Tabelle 8: Status von Wortlisteneinträgen

Status	Beschreibung
UNKNOWN	Unbekannt: Noch kein anderer Status vergeben.
EXCLUDE	Ausschließen: Im Index Wortstammreduktion weglassen, um den Term von anderen, inhaltlich verschiedenen Begriffen abzugrenzen, die nach der Reduktion fälschlicherweise denselben Wortstamm hätten (sinnvoll z. B. bei Kommune, Kommunion, Kommunismus, Kommunikation -> Wortstamm „kommun“).
ORDINARY	Gewöhnliches Wort: Ein ganz gewöhnliches Wort, kein Kandidat für das semantische Netz, d. h. kein Begriff der Fachsystematik.
IMPORTANT	Wichtig: Begriff aus der Fachsystematik, ein Kandidat für das semantische Netz.
STOPWORD	Stoppwort: Echte Stoppwörter wie „der“, „die“, „das“, „und“, die nicht in den Index aufgenommen werden sollen, weil sie nur ein „Rauschen“ in Dokumenten darstellen und so gut wie keine Deskriptivität besitzen, also keine Unterscheidbarkeit der Dokumente voneinander leisten.

Den Status UNKNOWN bekommen alle (neuen) Wörter, die noch keinen anderen Status besitzen und nicht standardmäßig aufgrund einer Plausibilitätsannahme der Liste ORDINARY (Zahlen und Wörter, in denen unter den ersten drei Zeichen ein Punkt oder ein Komma vorkommt) oder STOPWORD (Wörter bis einschließlich zwei Zeichen Länge) zugeschlagen werden.

Wie sich die einzelnen Status im System auswirken, zeigt die nachfolgende Übersicht.

Tabelle 9: Auswirkungen von Wortlisten-Status

Status	Unknown	Exclude	Ordinary	Important	Stopword
Bedeutung					

Im Index enthalten	Ja	Ja	Ja	Ja	Nein
Wortstammreduktion (für Suche im Index)	Ja	Nein	Ja	Ja	Nein
Ist Kandidat für semantisches Netz	Nein	Nein	Nein	Ja	Nein
Kann bei Suche gefunden werden	Ja	Ja	Ja	Ja	Nein
Wird bei Suche ignoriert (kann nicht gefunden werden)	Nein	Nein	Nein	Nein	Ja

Hinweis: die Terme mit Status UNKNOWN und ORDINARY werden vom System her im Hinblick auf die Suchfunktion gleich behandelt, es ist also kein Fehler in der Tabelle.

Die Bearbeitung der Wortlisten, d. h. vor allem die Zuweisung eines anderen Status zu einem Wort, erfolgt manuell über die Wortlisten-Maske (siehe Abbildung Abbildung 66). Es lassen sich mehrere Wörter selektieren und ihnen gemeinsam ein neuer Status vergeben. Da ein Großteil der Wörter gerade beim Aufbau einer Liste den Status UNKNOWN aufweist, gibt es im unteren Teil der Maske zwei Knöpfe, um alle aktuell dargestellten Terme mit ebendiesem Status ohne manuelle Einzelauswahl rasch als „Gewöhnliche Wörter“ bzw. als Stoppwörter zu kennzeichnen. Um eine versehentliche Betätigung dieser Funktionen zu verhindern, muss zuvor das Häkchen bei „Bestätigung“ neben dem jeweiligen Knopf gesetzt werden. Im Zusammenspiel der Filterfunktion und der Möglichkeit, viele Wörter auf einmal zu bearbeiten, erscheint der Aufwand für die manuelle Pflege der Listen vertretbar.

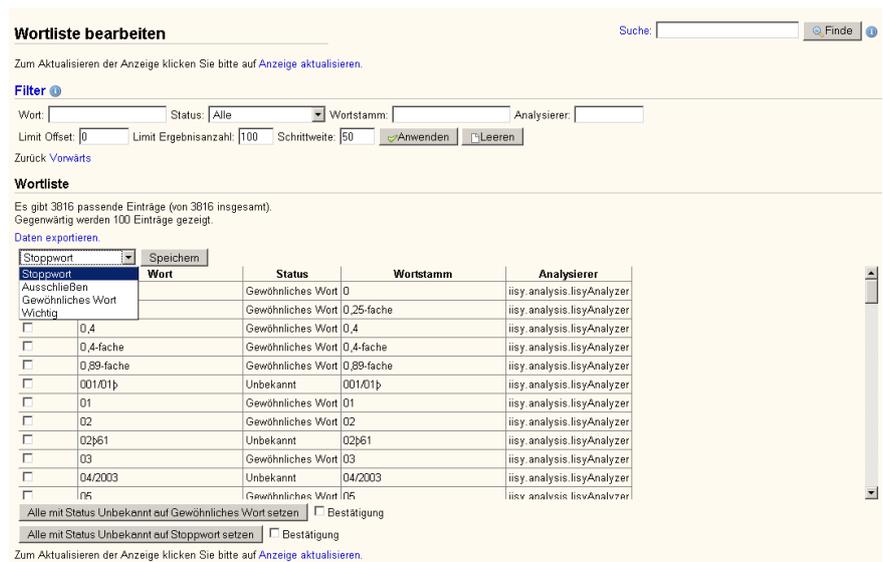


Abbildung 66: Bearbeitung von Wortlisten

In einer späteren Ausbaustufe sind für die automatische Statusvergabe für neu aufgenommene Terme noch weiter entwickelte Techniken wie Ähnlichkeitsvergleiche mit bereits enthaltenen Wörtern oder die Einbeziehung eines geeigneten Wörterbuches denkbar, um die Zahl der Terme, denen bisher nur der Status UNKNOWN zugewiesen werden kann, zu verringern.

Es sei an dieser Stelle darauf hingewiesen, dass das fortlaufende Mitschreiben von Wörtern während der Indexierung für Analyse- oder Lernzwecke einen stark negativen Effekt auf die Performance des Crawling-Vorganges hat. Der Crawling-Prozess weist eine deutlich längere Laufzeit auf! Außerdem erhöht sich der Hauptspeicherbedarf spürbar! Aus diesem Grunde empfiehlt es sich, das Mitschreiben von Wortlisten auszuschalten, falls diese nicht (mehr) benötigt werden. Ein dauerhafter Einsatz in Produktivsystemen ist nicht zu empfehlen. Stattdessen ist anzustreben, das Schreiben von Wortlisten nur gezielt einzusetzen und auf ausgewählte Datenquellen zu beschränken.

11.6.7 Lernen

Die Ausführungen an dieser Stelle beschränken sich auf das Lernen im semantischen Netz.

Das Lernen erfolgt für die Begriffe und die Beziehungen auf ähnliche Weise. Beide Objektklassen besitzen hierzu folgende Attribute:

- gültig
- Taxonomie
- akzeptiert
- löschar
- Treffer, Kurzzeit- und Langzeitrelevanz; jeweils mit Datum seit wann
- minimale und maximale Relevanz
- variable Relevanz

Beim Anlegen eines Eintrages werden die Attribute standardmäßig (abhängig davon, ob dies manuell oder automatisch geschieht) folgendermaßen gesetzt:

Tabelle 10: Attribute von Begriffen und Beziehungen im semantischen Netz

Attribut	Formel-Zeichen	manuell	automatisch
gültig	valid	Ja	nein
Taxonomie	taxo	nein	nein
akzeptiert	acpt	ja	nein
löschar	delable	nein	ja
Treffer	n_{hit}	0	Schwelle
Treffer, Zeitpunkt seit wann	t_{hc}	Zeitpunkt des Anlegens	
Kurzzeitrelevanz	r_{st}	0,5	0,5
Kurzzeitrelevanz, Zeitpunkt seit wann	t_{rst}	Zeitpunkt des Anlegens	
Langzeitrelevanz	r_{lt}	0,5	0,5
Langzeitrelevanz, Zeitpunkt seit wann	t_{rlt}	Zeitpunkt des Anlegens	
minimale Relevanz	r_{min}	0	0
maximale Relevanz	r_{max}	1	1
variable Relevanz	r_{var}	ja	ja

Beim automatischen Anlegen eines Eintrages wird bei n_{hit} eine negative Zahl eingetragen. Das vorrangige Ziel dieser Schwelle zu 0 ist es, falsch geschriebene Begriffe leichter ausfiltern zu können. Beispielsweise kann man den Wert auf -1 setzen. Möglicherweise wird aber die fehlerhafte Eingabe unmittelbar wiederholt. Um diesem Fall gerecht zu werden, nimmt man einen kleineren Wert. Ein zu kleiner Wert sollte jedoch nicht gewählt werden, da ansonsten selten gesuchte Begriffe aufgrund der hohen Schwelle nicht automatisch gelernt werden können. Über die genannten Attribute kann das Lernverhalten der einzelnen Elemente gesteuert bzw. beeinflusst werden.

Obwohl das Lernen im semantischen Netz prinzipiell automatisch ablaufen könnte, ist der Rückgriff auf das Wissen des menschlichen Experten und dem ihm verfügbaren umfangreichen Kontext der realen Welt trotzdem wichtig.

So macht es durchaus Sinn, nicht nur die aufrufbaren Lernfunktionen zu verwenden, sondern mit Hilfe der Filtermöglichkeiten bei den Begriffs- und Beziehungsfunktionen sich bestimmte Elemente anzusehen und jeweils manuell vorzubereiten. Da die Lernfunktionen auch ein stufenweises Vergessen des Kurzzeitgedächtnisses beinhalten, ist es effektiv und besonders effizient, manuelles Bearbeiten einerseits (siehe 11.5.1) und Lernfunktion aufzurufen andererseits im Wechsel durchzuführen. Die Funktionen, die für das Lernen zur Verfügung stehen, sind aus Abbildung 67 ersichtlich.

Lernen im semantischen Netz

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Lernen im semantischen Netz

Bereich	Eigenschaft	min	max	Begriffe	Beziehungen	Korrektur
Bedingungen	Gültige Einträge	64	-	14034 ✓	3674 ✓	
	Gültige Einträge mit Treffern	8	-	870 ✓	836 ✓	
	Reife Einträge mit Treffern (Reifungstage: 4)	8	-	870 ✓	836 ✓	
	Gesamte Treffer über alle gültigen Einträge	16	-	3746 ✓	2995 ✓	
	Gesamte Treffer über alle reifen Einträge	16	-	3746 ✓	2995 ✓	
Fehler	Relevanz: Langzeit: Datum fehlt	-	0	0 ✓	0 ✓	(nicht erforderlich)
	Relevanz: Kurzzeit: Datum fehlt	-	0	0 ✓	0 ✓	(nicht erforderlich)
	Relevanz: Zähler-Rückstellung: Datum fehlt	-	0	0 ✓	0 ✓	(nicht erforderlich)
	Relevanz: Folge: Langzeit - Kurzzeit	-	0	0 ✓	0 ✓	(nicht erforderlich)
	Relevanz: Folge: Kurzzeit - Zähler-Rückstellung	-	0	0 ✓	0 ✓	(nicht erforderlich)
Aktionen	Funktion					
	Lernen			<input type="button" value="Ausführen"/>	<input type="button" value="Ausführen"/>	
	Reife Einträge mit negativen Treffern löschen			<input type="button" value="Ausführen"/>	<input type="button" value="Ausführen"/>	
	Einträge mit negativen Treffern löschen			<input type="button" value="Ausführen"/>	<input type="button" value="Ausführen"/>	
	Treffer und Daten (Mz. von Datum) komplett zurücksetzen			<input type="button" value="Ausführen"/>	<input type="button" value="Ausführen"/>	

Zum Aktualisieren der Anzeige klicken Sie bitte auf [Anzeige aktualisieren](#).

Abbildung 67: Lernen im semantischen Netz

Wie man aus der ersten Spalte mit der Überschrift „Bereich“ erkennen kann, gibt es einerseits Lernvoraussetzungen und andererseits Lernaktionen. Das Lernen erfolgt an dieser Stelle für die Begriffe und Beziehungen unabhängig voneinander, aber (wie man an den entsprechenden beiden Spalten sieht) in analoger Weise.

Bedingungen

Über die Bedingungen ist im Wesentlichen festgelegt, ab wann ein Lernen überhaupt erst Sinn macht:

- Gültige Einträge:
Das sind diejenigen Einträge, bei denen „valid“ wahr ist
- Gültige Einträge mit Treffern:
Einträge, bei denen zusätzlich n_{hit} einen positiven Wert aufweist
- Reife Einträge mit Treffern:
Hier muss zusätzlich das Datum in t_{hc} mindestens vier Tage zurückliegen
- Gesamte Treffer über alle gültigen bzw. reifen Einträge:
Hier werden nicht die Einträge gezählt, die Treffer enthalten, sondern die Gesamtzahl der Treffer

Die gewählten Minimalwerte sind Annahmen, die erst durch den Einsatz in der Praxis bestätigt bzw. korrigiert werden können. Man kann aber prinzipiell davon ausgehen, dass bessere Ergebnisse erzielt werden, wenn die tatsächlichen Werte größer sind. Andererseits bedeuten größere Werte auch zeitlich größere Lernintervalle.

Fehler

Es gibt zwei prinzipielle Fehler:

- Relevanz: Datum fehlt
(Langzeit, Kurzzeit bzw. Zähler-Rückstellung)
- Relevanz: zeitliche Folge
(Langzeit – Kurzzeit bzw. Kurzzeit - Zähler-Rückstellung)

Diese Fehler können sowohl für jeden Eintrag separat oder – wesentlich einfacher – für einen bestimmten Fehler über alle Einträge bei den Begriffen und Beziehungen auf einmal behoben werden. Falls ein bestimmter Fehler existiert, dann wird in der Spalte „Korrektur“ eine Schaltfläche mit der Aufschrift „Ausführen“ angeboten, ansonsten steht hier wie dargestellt „(nicht erforderlich)“.

Aktionen

Für das Lernen werden diese Aktionen angeboten:

- Lernen
- Reife Einträge mit negativen Treffern löschen
- Einträge mit negativen Treffern löschen
- Treffer und Daten (Mz. von Datum) komplett zurücksetzen

Sobald alle Bedingungen erfüllt und alle Fehler behoben sind, werden in der Zeile „Lernen“ in den Spalten „Begriffe“ und „Beziehungen“ Schaltfläche mit der Aufschrift „Ausführen“ angeboten, ansonsten steht hier „(nicht möglich)“. Damit IRIS durch Neuberechnung der Relevanzen lernt, genügt es, die entsprechenden Schaltflächen zu betätigen.

Alle anderen Aktionen können immer ausgeführt werden (also unabhängig davon, ob die Voraussetzungen erfüllt sind oder nicht). Diese Funktionen erfüllen Teile einer wichtigen Nebenfunktion des Lernens, nämlich die des Vergessens.

Sanfte Bereinigung des Kurzzeitgedächtnisses³⁵:

Das Löschen der reifen Einträge mit negativen Treffern entfernt alle Einträge, die nur selten gesucht werden oder bei der Suche einfach nur falsch geschrieben wurden. Da aber nicht ausgeschlossen werden kann, dass es sich im Einzelfall um einen wichtigen Eintrag handelt, ist es durchaus sinnvoll, vorab alle entsprechenden Einträge zumindest kurz zu sichten und gegebenenfalls zu bearbeiten.

Massive Bereinigung des Kurzzeitgedächtnisses:

Das Löschen der Einträge mit negativen Treffern entfernt darüber hinaus auch die Einträge, die noch nicht die Reifezeit erreicht haben.

Treffer und Daten (Mz. von Datum) komplett zurücksetzen:

Damit lässt sich das semantische Netz teilweise zurücksetzen. Die aktuellen Werte der Langzeit- und Kurzzeitrelevanzen bleiben dabei aber zunächst erhalten. Eine Auswirkung ergibt sich erst beim nächsten Lernen. In der Regel ist es sinnvoll, zuvor das Kurzzeitgedächtnis mit einer der beiden zuvor aufgeführten Funktionen zu bereinigen.

11.6.8 Verwaltung automatisierter Aufgaben

Die Verwendung der in IRIS so genannten „Automatisierten Aufgaben“ wurde – da diese logisch zusammenhängen – bereits in Verbindung mit der Indexierung erklärt. Es sei daher an dieser Stelle lediglich auf das Kapitel 11.6.4 und dort insbesondere auf das Unterkapitel 11.6.4.2.2 verwiesen.

11.7 Erweiterungen von Regain

Als Grundphilosophie wurde bei den Erweiterungen Wert darauf gelegt, den bestehenden Quellcode so weit als möglich im Originalzustand zu erhalten und die notwendigen Eingriffe auf ein Minimum zu reduzieren. Die zusätzlichen Fähigkeiten wurden wenn möglich als neue Pakete bzw. Quelldateien

³⁵ Diese Aktion wird auch beim eigentlichen Lernen als erster Schritt ausgeführt.

dem Projekt hinzugefügt. Direkte Umbauten im Quellcode waren dennoch erforderlich, um die IRIS-spezifischen Modifikationen zu aktivieren. Außerdem wurde darauf geachtet, dass „Regain“ trotz der Änderungen auch weiterhin in seinem ursprünglichen Funktionsumfang benutzt werden kann. Diese Grundhaltung wurde im Übrigen auch für die Anpassungen der anderen im Gesamtsystem „IRIS“ eingesetzten Fremdsoftware „JSPWiki“, „JdbcProvider“ und „Quartz“ angewandt.

Analog zum Vorgehen beim Quellcode selbst wurden die neuen Funktionen als optionale Erweiterungen des Formates der IRIS-Indexierungskonfiguration realisiert. Das heißt, die zusätzlichen Funktionen werden durch spezielle Einstellungen in den Crawler-Konfigurationsdateien aktiviert, die vom ursprünglichen „Regain“ einfach ignoriert werden, da diese dort unbekannt sind.

Trotz größter Sorgfalt ist aber dennoch niemals ganz auszuschließen, dass die vorgenommenen Modifikationen in bestimmten Konstellationen mit dem Original kollidieren, sei es hinsichtlich einer Konfigurationseinstellung, sei es hinsichtlich einer Passage im Quellcode.

Bezüglich der im Original enthaltenen Einstellungsmöglichkeiten von „Regain“ sei an dieser Stelle auf das Anwenderhandbuch ([Schn3]) sowie auf die Hilfe-seite von Regain ([Regain]) verwiesen. Die bei IRIS neu eingeführten Elemente werden hier erklärt. Es sei angemerkt, dass die nachfolgend in diesem Kapitel dargelegten Erläuterungen stark ins Detail gehen und dem Leser, der nicht konkret mit dem System arbeitet bzw. administriert, zu speziell sein mögen.

Für alle Erklärungen gilt, dass sich die genannten XML-Elemente (oder englisch „Tags“: etwa Kennzeichnung oder Marke) innerhalb des Hauptelements `<configuration>` der Konfigurationsdatei befinden. Dieses ist also gemeinsamer Bezugspunkt für alle nachfolgenden Ausführungen.

Dem Element `<start>` für die Startliste wurden die beiden zusätzlichen optionalen Attribute `username` (Benutzername) und `password` (Passwort) hinzugefügt. Damit wird veranlasst, dass der Crawler eine authentifizierte Verbindung mit dem angegebenen Ziel und allen gefundenen Unterseiten unter Verwendung des genannten Nutzernamens und Passwortes aufbaut. Der Gebrauch dieser Erweiterung macht freilich nur bei geschützten und für IRIS ausdrücklich zur Indexierung erlaubten Webauftritten – oder Teilen eines solchen – Sinn.

Eine umfangreichere und im Umfeld von IRIS für die Praxis bedeutsamere Erweiterung betrifft ebenfalls das Element `<start>`. Statt der Angabe einer URL kann mit Hilfe des neu eingeführten Elements `<source>` eine alternative Datenquelle angegeben werden. `<source>` erwartet das Attribut `type` zur Festlegung der Art der Datenquelle. Gegenwärtig unterstützt werden die Typen `db` bzw. `database` für eine Datenbank und `xml` für eine XML-Datei mit Metainformationen über Dokumente. Mit Hilfe von Subelementen wird die Datenquelle konfiguriert. Welche Elemente das sind, hängt vom vorgenannten Typ der Datenquelle ab.

Im Falle einer Datenbank-Quelle (also Typ `db` oder `database`) werden die nachfolgend erklärten Elemente `<url>`, `<drivername>`, `<username>`, `<password>` und `<sql>` benötigt. Der Körper des `<url>`-Tags definiert den Zugangs-String, mit dem die Datenbank angesprochen wird. Er setzt sich zusammen aus dem obligatorischen Präfix `db://` (zur Unterscheidung von anderen Protokollen wie `file://` oder `http://`), der Datenbankanbieter-Spezifikationskennung (im Falle einer Mysql-Datenbank z. B. `jdbc:mysql:`), dem Host der Datenbank (bei einer lokal verfügbaren Datenbank z. B.

//127.0.0.1/) und dem Namen der konkret anzusprechenden Datenbank (die IRIS-Datenbank heißt z. B. **iisy**). Der gesamte `<url>`-Tag könnte komplett also beispielsweise `<url>db://jdbc:mysql://127.0.0.1/iisy</url>` lauten. Mit dem Tag `<drivername>` wird festgelegt, welcher Datenbank-Treiber konkret benutzt werden soll. Gegenwärtig ist nur der Standard-Treiber von Mysql in IRIS eingebunden. Seine Klassenbezeichnung ist `com.mysql.jdbc.Driver`. Treiber für Java sind für viele andere Datenbanken verfügbar, müssen vor einer Verwendung aber erst in die Applikation eingebunden werden. In den Tags `<username>` bzw. `<password>` werden der Nutzernamen bzw. das Passwort festgelegt, mit denen sich der Crawler bei der Datenbank anmelden soll. Im Element `<sql>` wird schließlich die eigentliche Abfrage auf den Datenbestand in Form einer SQL-Anweisung festgelegt. Das Ergebnis wird von der Datenbank zeilenweise entgegengenommen und ausgewertet. Die Übersetzung des Ergebnisses in eine für den Crawler verständliche Form, d. h. die Angaben zur Interpretation der Rückgabewerte, wird weiter unten bei den Feld-Elementen erklärt.

Der zweite derzeit implementierte Typ von alternativen Datenquellen ist die XML-Metadaten-Datei (Typ `xml`). Hierbei gibt es die spezifischen Elemente `<file>` und `<docTag>`. Ersteres legt die konkret auszulesende XML-Datei fest, Letzteres die Bezeichnung des Tags, das die einzelnen Dokumentenbeschreibungen in dieser Datei umschließt. Dabei ist zu beachten, dass die XML-Quelle folgenden grundsätzlichen Aufbau haben muss: die Tags mit den Metadaten zu Dokumenten haben alle dieselbe Bezeichnung und sind gleichrangig innerhalb eines alle umschließenden Datei-Oberelementes hintereinander gereiht. Beim Auslesen wird die ganze Datei durchlaufen. Alle gefundenen Tags mit der in `<docTag>` genannten Bezeichnung (genauer gesagt: deren sämtliche Subelemente) werden analog der zeilenweisen Verarbeitung bei der Datenbankabfrage vom Crawler mit Hilfe der weiter unten noch erklärten Feld-Elemente einzeln ausgewertet.

Die zuvor bereits erwähnten Feld-Elemente sind allen alternativen, also mit Hilfe des neu eingeführten `<source>`-Tags definierten, Datenquellen gemein. Mit ihrer Hilfe wird ein Mapping, also eine Abbildung, der konkret von der jeweiligen Datenquelle gelieferten Ergebnisse in eine einheitliche und vom Crawler interpretierbare Form realisiert. Dazu dient das Element `<field>`. Es erwartet die Pflichtattribute `column`, `fieldname`, `store` und `index`, optional bei Bedarf zusätzlich `prepare` und `prepareContentType`. Mit `column` wird der Name der Spalte aus der Ergebnismenge einer Datenbankabfrage (bzw. im Falle einer XML-Metadaten-Quelle das Element) bezeichnet, für das die Abbildung gelten soll. Das Attribut `fieldname` gibt an, in welches Feld in Lucene der Inhalt geschrieben werden soll. In „Regain“ gibt es mehrere fest vergebene Feldnamen, die jeweils eine bestimmte Funktion haben (siehe nachfolgende Übersicht). Man ist jedoch nicht auf diesen Katalog von vorgegebenen Feldnamen beschränkt.

Tabelle 11: Von Regain standardmäßig genutzte Felder in Lucene

Feldname	Bedeutung	Pflichtfeld
url	URL des Dokumentes	Ja
title	Titel des Dokumentes	Nein
summary	Kurze Zusammenfassung	Nein
headlines	Herausragende Schlagzeilen	Nein
size	Größe des Dokumentes	Nein

content	Text-Inhalt	Ja
last-modified	Zeitpunkt der letzten Änderung	Ja
path	Pfad zum Dokument	Nein
preparation-error	Marker für fehlerhaften Eintrag	Nein*
groups	Nutzer-Gruppen-Einteilung	Nein

*) Das Feld wird bei Bedarf vom Crawler selbst vergeben.

Wie Lucene mit den Inhalten des aktuellen Feldes verfährt, wird mit den Attributen `store` und `index` festgelegt. Dazu sollte man wissen, was die beiden Einstellungen für Auswirkungen haben. Der Store-Wert bestimmt, ob bzw. wie der Feld-Inhalt in seiner originalen Form im Index abgespeichert wird. Beim Auslesen des Index wird der Inhalt aller abgespeicherten Felder dann stets für jeden Treffer in das entsprechende Java-Objekt in den Hauptspeicher geladen. Daten, die nicht in ihrer ursprünglichen Form benötigt werden, sollten daher nicht extra abgespeichert werden. Die möglichen Einstellungen zur Abspeicherung zeigt die nachfolgende Übersicht.

Tabelle 12: Einstellungsmöglichkeiten zum Speicherungsverhalten von Feldern in Lucene

Einstellung für Store	Bedeutung
COMPRESS	Original Feld-Inhalt in komprimierter Form im Index abspeichern
NO	Original Feld-Inhalt nicht abspeichern
YES	Original Feld-Inhalt im Index abspeichern

Unabhängig von der Einstellung zur Abspeicherung des Inhalts eines Feldes ist die Festlegung bezüglich der Indexierung desselben. Indexierung bedeutet, dass für die einzelnen Terme des Inhalts Referenzen auf das jeweilige Feld im Index angelegt werden. Die möglichen Werte zeigt nachfolgende Übersicht.

Tabelle 13: Einstellungsmöglichkeiten zum Indexierungsverhalten von Feldern in Lucene

Einstellung für Index	Bedeutung
NO	Feld-Inhalt nicht indexieren
NO_NORMS	Feld-Inhalt ohne Text-Analysierer indexieren, sodass er gefunden werden kann, jedoch zur Speicherersparnis ohne Feldlängen-Normalisierungs-Informationen
TOKENIZED	Feld-Inhalt unter Verwendung eines Text-Analysierers indexieren, sodass er gefunden werden kann
UN_TOKENIZED	Feld-Inhalt ohne Text-Analysierer (also als ein einzelner Term) indexieren, sodass er gefunden werden kann

Die Kombination aus `store` und `index` legt also fest, ob ein Feldinhalt ausgelesen werden kann (Abspeicherung) und ob ein Feldinhalt (bzw. dessen Terme) bei einer Suche gefunden werden kann (Indexierung). Zumindest die „Regain“-Pflichtfelder müssen beim Mapping berücksichtigt werden. Speziell für das Feld „last-modified“ ist das vorgegebene Datumsformat zu beachten.

Das optionale und in der Praxis wohl eher selten benötigte Attribut `prepare` ermöglicht es, den von der Datenquelle empfangenen Inhalt vor der üblichen Weiterverarbeitung einem Präparator zuzuführen. Das kann vor allem bei binären Dateiinhalten aus einer Datenbank sinnvoll sein. Wird z. B. der in einer Datenbank abgelegte binäre Dateiinhalt etwa einer PDF-Datei abgerufen, sollte nicht dieser Byte-Strom als Inhalt gelten, sondern vielmehr muss erst vom üblicherweise dafür zuständigen PdfPreparator der Klartext als Inhalt extrahiert werden. Denn normalerweise ist man in IRIS an den textlichen Inhalten interessiert. Dieser Vor-Verarbeitungsschritt für Inhalte aus einer alternativen Datenquelle kann gezielt aktiviert werden, wenn das optionale Attribut `prepare` auf **yes** oder **true** gesetzt wird. Im Zusammenhang damit ist auch das zweite optionale Attribut `prepareContentType` zu sehen und ist nur gültig, wenn auch `prepare` aktiv ist. Es kommt zum Zuge, falls aus dem übermittelten Wert für die URL nicht hervorgeht, welcher Präparator für den oben genannten Schritt zuständig wäre. Für das oben genannte Beispiel könnte als Wert also **pdf** in Frage kommen, wenn die mit übermittelte URL keinen Hinweis auf den echten Dateinamen aufweist, sondern etwa nur eine Identifikationsnummer oder dergleichen.

Soweit die Erweiterungen des Elementes `<start>`. Die genaue Benutzung mag beim Durchsehen von echten Beispielen klarer werden. Hierzu seien dem Anwender die in IRIS beigefügten Vorlagen für Indexierungskonfigurationen nahe gelegt (vgl. Kapitel 11.6.4.1).

Für die zuvor genannten alternativen Datenquellen (Datenbank bzw. XML-Metadatenfile) wurden die beiden „Protokoll“-Bezeichnungen „db://“ bzw. „xml://“ eingeführt, um sie von anderen, herkömmlichen Datenquellen (Webseiten bzw. lokale Dateien mit den Präfixen „http://“ bzw. „file://“) zu unterscheiden. Damit Regain sie auch als gültig akzeptiert, müssen sie wie die anderen in der Weißen Liste vermerkt werden. Das heißt konkret, dass im Konfigurationsabschnitt der Weißen Liste für die Benutzung solcher alternativen Datenquellen jeweils ein Eintrag wie `<prefix>db://</prefix>` eingetragen werden muss.

Ganz neu im Zusammenhang mit IRIS wurde das Element `<documents>` eingeführt. Seine Einstellungen werden nur bei Verwendung eines Datenbank-Schreibers mit dem ebenfalls neu eingeführten Tag `<writer>` aktiv, welches weiter unten erläutert wird. Das Tag `<documents>` dient der einfachen Zuordnung von Metadaten zu den beim IRIS-Crawling-Vorgang Schritt 1 (vgl. Kapitel 10.4.5) aufgefundenen Dokumenten. Das setzt natürlich voraus, dass der Redakteur einigermaßen darüber Bescheid weiß, welche Dokumente mit Hilfe der aktuellen Konfigurationsdatei durchlaufen werden.

Die Dokumenten-Einträge in der IRIS-Datenbank haben zwei Angaben darüber, welcher Kategorie das Dokument zugeordnet ist (Spalte `Categories`) und welche Themengebiete im Dokument behandelt werden (Spalte `Topics`). Diese Angaben können in der Dokumentenverwaltung (siehe Kapitel 11.6.5) für die Einträge einzeln bearbeitet werden. Um den mühsamen Vorgang zu beschleunigen, können den Dokumenten schon beim Crawling Vorgabewerte zugeteilt werden. Dies geschieht innerhalb des Elementes `<documents>` mit den Tags `<defaultCategories>` für Kategorien bzw. `<defaultTopics>` für Themengebiete. Innerhalb dieser Elemente wird jeweils eine Liste von Tags `<category>` bzw. `<topic>` aufgeführt, in deren Körper der zu vergebende Wert steht. Beispielsweise würde mit folgendem Auszug aus einer Konfigurationsdatei erreicht, dass alle gefundenen Dokumente der Kategorie „Science“ und den Themengebieten „Holz“ sowie „Holzbau“ zugeordnet würden:

```

<documents>
  <defaultCategories>
    <category>Science</category>
  </defaultCategories>
  <defaultTopics>
    <topic>Holz</topic>
    <topic>Holzbau</topic>
  </defaultTopics>
</documents>

```

Dieses global für alle mit Hilfe einer bestimmten Konfigurationsdatei gefundenen Dokumente wirkende Verfahren ist jedoch nicht immer ausreichend. Es wurde daher mit dem Tag `<correlation>` (Zuordnung) ein weiteres Element eingeführt, um allgemein Gruppen von Dokumenten schon bei der Aufnahme in die Datenbank gezielt Metadaten/Eigenschaftswerte zuweisen zu können. Das Element `<correlation>` erwartet die Pflichtattribute `name`, `order` sowie entweder `prefix` oder `regex`. Das Attribut `name` dient der Gruppierung der Zuordnungen selbst. Alle Zuordnungen mit demselben Namen gehören zusammen, werden also gemeinsam behandelt. Derzeit in IRIS genutzt werden **access** (für Zugriffsrechte), **categories** (für Kategorien) und **topics** (für Themengebiete). Mit dem Attribut `order` wird die Rangfolge für den späteren Abgleich mit den zu prüfenden URLs festgelegt. Als Werte sind hier ganze, nicht-negative Zahlen zulässig. Die Reihenfolge ist aufsteigend, d. h. die Zuordnung mit dem kleineren Wert für `order` hat Vorrang. Alle Zuordnungen einer Gruppe (siehe Attribut `name`) müssen unterschiedliche Werte für die Rangfolge haben, damit die Prioritätsregelung eindeutig ist. Ein Verstoß gegen diese Regel wird beim Einlesen der Konfigurationsdatei erkannt und mit einer entsprechenden Fehlermeldung quittiert. Das dritte Pflichtattribut ist entweder `prefix` oder `regex` und legt fest, für welche Dokumenten-URLs die Zuordnung gelten soll. Im Falle von `prefix` sind dies alle URLs, die mit dem genannten Präfix beginnen. Bei `regex` ist als Wert ein regulärer Ausdruck anzugeben. Alle URLs, die diesen regulären Ausdruck erfüllen, gelten als zutreffend.

Mit den oben genannten Attributen werden also die Gruppierung der Zuordnungen, deren Priorisierung innerhalb einer Gruppe und die URLs festgelegt, für die diese gelten sollen. Welche Eigenschaften den Dokumenten genau zugeordnet werden sollen, wird in den möglichen Sub-Elementen `<value>` oder `<property>` definiert. Mit jedem Tag `<value>` kann in seinem Körper ein Wert angegeben werden (analog etwa dem Tag `<topic>` im Element `<defaultTopics>`). Alle diese Werte sind gleichrangig und nicht unterscheidbar. Beim Tag `<property>` hingegen werden mit Hilfe der beiden Pflichtattribute `key` und `value` identifizierbare Schlüssel-Wert-Paare definiert. Derzeit werden in der Gruppe **access** die Schlüssel **source** (für die Quelle) bzw. **userlevel** (für den Dokumenten-Nutzerlevel) unterstützt. In der Gruppe **access** können somit die Zugriffsrechte auf die Dokumente festgelegt werden (vgl. hierzu auch Kapitel 11.6.5).

Zum besseren Verständnis sei noch einmal kurz die Wirkung an folgendem Beispiel einer Zugriffsberechtigung gemäß der in IRIS vorgesehenen Abstufung erläutert. Zunächst der entsprechende Ausschnitt der Konfigurationsdatei:

```

<documents>
  <correlation name="access" order="0" prefix="file:///data/docs/private">
    <property key="source" value="www.partner.test.de/order" />
    <property key="userlevel" value="2" />
  </correlation>

  <correlation name="access" order="1" prefix="file:///data/docs/registered">
    <property key="source" value="www.partner.test.de/" />
    <property key="userlevel" value="1" />
  </correlation>

```

```

<correlation name="access" order="2" prefix="file:///data/docs/">
  <property key="source" value="www.partner.test.de/public" />
  <property key="userlevel" value="0" />
</correlation>
</documents>

```

Die drei aufgeführten Zuordnungen gehören alle zur Gruppe **access**. Wird nun vom IRIS-Crawler ein Dokument verarbeitet, wird anhand von dessen URL geprüft, ob eine der Zuordnungen zutrifft. Dazu werden gemäß der mittels der Attribute `order` festgelegten Reihenfolge die Zuordnungen durchlaufen und jeweils geprüft, ob die URL mit dem gegebenen Präfix beginnt. Ist das bei einer Zuordnung der Fall, gilt sie (weitere, ebenfalls zutreffende Zuordnungen mit niedrigerer Priorität werden ignoriert). Da es sich um die Gruppe **access** handelt, sucht der IRIS-Crawler nach den benannten Schlüsseln **source** bzw. **userlevel**. Die Werte, die dort hinterlegt sind, werden beim Dokument als Quelle bzw. als Benutzer-Level vermerkt.

Im Zusammenhang mit dem soeben erläuterten Element `<documents>` wurde bereits das Tag `<writer>` erwähnt. Mit ihm kann statt einem Lucene-Index ein alternatives Datenziel für die Dokumenteneinträge festgelegt werden. Das wird beim IRIS Crawling Schritt 1 benutzt (vgl. Kapitel 10.4.5), wo als Ziel die IRIS-Datenbank anzugeben ist. `<writer>` erwartet das Attribut `type` zur Festlegung des Typs des Schreibers. Bisher gibt es nur einen Datenbank-Schreiber, weswegen hier als Wert stets **database** anzugeben ist. Es gibt derzeit keine Pläne, einen anderen Schreiber zu implementieren.

Im Falle eines Datenbank-Schreibers ist die Datenbank mit den Tags `<drvClass>` (für den Java-Datenbanktreiber), `<vendorSpec>` (Datenbankanbieter-Spezifikationskennung), `<hostname>` (Host der Datenbank), `<name>` (Name der Datenbank) `<user>` (Nutzername) und `<pwd>` (Passwort) zu beschreiben. Die Aufteilung wurde hier im Vergleich zu den entsprechenden Angaben für eine Datenbank-Quelle in der Startliste (siehe weiter oben) etwas entzerrt. Im Prinzip geht es aber analog um dieselben Daten.

Mit dem optionalen Unter-Element `<wordlist>` kann im Falle eines Datenbank-Schreibers festgelegt werden, ob bzw. mit welcher Strategie während des IRIS-Crawling-Vorganges Wortlisten mitgeschrieben werden sollen (fehlt das Tag, werden keine Wortlisten geschrieben). Dazu erwartet `<wordlist>` das Attribut `strategy`. Mit den Werten **false**, **no** oder **off** wird das Schreiben ausdrücklich abgeschaltet.

Grundsätzlich sei an dieser Stelle wie auch schon in Kapitel 11.6.6 nochmals darauf hingewiesen, dass das Mitschreiben von Wortlisten in allen Varianten generell stark negative Effekte auf Performance und Hauptspeicherverbrauch während des Crawling-Vorganges mit sich bringt. Sollen dennoch Wortlisten erstellt werden, gibt es mehrere Varianten, die je nach Umständen schneller oder langsamer sein können. Hier wurde versucht, die negativen Effekte mit verschiedenen Mitteln zu reduzieren. Ein abschließendes Ergebnis, welche Variante in welcher Ausgestaltung wann am besten geeignet ist, liegt bislang nicht vor.

Mit `list` wird bestimmt, dass die Terme eines Dokumentes jeweils in einer Datenstruktur vom Typ `LinkedList` (einfach verkettete Liste) gesammelt werden, um sie dann der Wortliste in der Datenbank hinzuzufügen. Der Wert `set` bestimmt, statt einer Liste eine Mengenstruktur (`Set`) zu benutzen (d. h. gleiche Terme werden nicht mehrfach im Speicher gehalten; es bedeutet aber mehr Rechenaufwand, um dies zu erreichen).

Bei Mengen kann ein eigener Cache benutzt werden, um die Anzahl der unnötigen Datenbank-Fehlzugriffe bei schon in der Wortliste vorhandenen Termen zu reduzieren. Er arbeitet so, dass die Menge der im Cache enthaltenen Terme von der Menge der in einem Dokument gefundenen Terme entfernt wird (also mengenmäßige Differenzbildung), bevor versucht wird, sie in die Datenbank zu schreiben. Die (Anfangs-)Größe wird als Anzahl von Termen unmittelbar nach dem Schlüsselwort **set** angefügt (also z. B. **set0** um keinen Cache zu benutzen oder **set2000** für einen Cache von bis zu 2000 Termen). Wird ein Cache benutzt, wird er zu Beginn bis maximal zur angegebenen Größe mit den kürzesten bereits in der Datenbank in der Wortliste vorhandenen Termen gefüllt. Der Cache kann noch anwachsen, wenn das Suffix **growing** angefügt wird (z. B. legt **set5000growing** die Benutzung eines Caches von Anfangs bis zu 5000 Termen aus der Datenbank fest, wobei der Cache aber um alle neu in Dokumenten gefundenen Terme anwachsen wird). Ein anwachsender Cache ist insofern gefährlich, als sein Speicherverbrauch letztlich unbegrenzt anwachsen und auf Dauer damit sogar den Crawler zum Absturz bringen kann. Außerdem steigt der nötige Rechenaufwand für die Mengenoperation immer weiter an. Bei Testläufen stieg die Zahl der im Speicher gehaltenen Terme abhängig von Anzahl und Inhalt der zugrunde gelegten Dokumente zumeist rasch auf einige Zehntausend und mehr an, wobei die negativen Effekte spürbar wurden und vereinzelt Abstürze auftraten. Von der Verwendung der Einstellung **growing** wird daher eher abgeraten.

Zur Benutzung des `<writer>`-Tags sind noch zwei wichtige Hinweise anzuführen:

- Die Verwendung eines alternativen Schreibers deaktiviert die Angabe zum Speicherort des Lucene-Index im `<dir>`-Tag des `<searchIndex>`-Elements.
- Das Schreiben von Wortlisten funktioniert nur bei Verwendung eines auf dem `lisyAnalyzer` basierenden Analysierers, also mit einem für IRIS angepassten Analysierer.

Die Anwendung all der neuen Konfigurationsmöglichkeiten mag aus der Anschauung von Vorlagen und Beispielen und durch Ausprobieren an Test-Konfigurationen und der Beobachtung der Auswirkungen für den Administrator eines IRIS-Systems klarer werden.

12 Ausblick

Die Software-Entwicklung wurde im ersten Quartal 2008 im Wesentlichen abgeschlossen. Seither werden daran vor allem Verbesserungen, Fehlerbehebungen und kleinere Funktionserweiterungen zur Abrundung des Systems vorgenommen. Aus Sicht des Projektteams werden die wesentlichen Forderungen, die im Rahmen dieses Forschungsprojektes an das System gestellt werden, bewältigt. Gleichwohl ist es weiter ausbaufähig.

Seit Mitte August 2007 ist IRIS als Testversion im Internet unter der Adresse <http://iris.fh-rosenheim.de> verfügbar. Seither werden Praxiserfahrungen im Betrieb gesammelt, z. B. bzgl. des Einspielens von Patches und Updates. Aufgrund der bedauerlicherweise zuletzt noch immer verhaltenen Resonanz seitens der Zielgruppen konnten jedoch noch kaum Erkenntnisse hinsichtlich der Effektivität des „Lernenden Systems“ – welches auf ein gewisses Mindestmaß an Nutzerinteraktion angewiesen ist – oder der Bewältigung von Spitzenlasten gewonnen werden.

Um IRIS nach dem Ende der Projektlaufzeit in die Praxis zu überführen, wurde im März 2008 beim Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) im Rahmen des Programms „FIT für den Wissenswettbewerb – Transferaktivitäten“ unter dem Titel „Ontologisch-semantisches integriertes Relationales Informationssystem – OSIRIS“ eine Projektskizze eingereicht. Der Schwerpunkt dieses anvisierten Nachfolgeprojektes ist der Transfer der bisher geleisteten Arbeiten und Ergebnisse in die Praxis. Zum Zeitpunkt der Erstellung dieses Berichtes gab es von Seiten des DLR noch keine Stellungnahme zu einer möglichen Förderung von OSIRIS.

13 Glossar

Alle im folgenden Glossar aufgeführten Definitionen entstammen, soweit keine andere Quellenangabe aufgeführt ist, aus der Online Enzyklopädie Wikipedia oder sind selbst verfasst. Alle Abrufe wurden am 12.06.06 durchgeführt.

Datenbank

In der Informatik eine Software (Datenbankmanagementsystem DBMS) zur strukturierten Speicherung und Verwaltung großer Datenmengen. Die Software operiert auf den Daten (die eigentliche „Datenbank“) und stellt geeignete Funktionen zur Abfrage sowie zum Hinzufügen, Ändern und Löschen von Datensätzen zur Verfügung. Hierbei hat sich SQL als Standard etabliert.

Im weiteren Sinne wird in diesem Dokument der Begriff manchmal allgemein als Sammlung von Dokumenten und Informationen benutzt und mit dem entsprechenden Internetauftritt eines Anbieters gleichgesetzt.

Index

Hier: Geordnete Datenstruktur mit Begriffen und den zugehörigen Fundstellen aus den verfügbaren Dokumenten.

Information

Information (von lateinisch: informare 'bilden, durch Unterweisung Gestalt geben') ist ein potenziell oder tatsächlich vorhandenes nutzbares oder genutztes Muster von Materie und/oder Energieformen, die für einen Betrachter innerhalb eines bestimmten Kontextes relevant ist. Wesentlich für die Information ist die Wiedererkennbarkeit sowie der Neuigkeitsgehalt. Das verwendete Muster verändert den Zustand eines Betrachters – im menschlichen Zusammenhang insbesondere dessen Wissen. Formaler ist Information die Beseitigung von Unbestimmtheit.

Informationssystem

dient der rechnergestützten Erfassung, Speicherung, Verarbeitung, Pflege, Analyse, Benutzung, Verbreitung, Disposition, Übertragung und Anzeige von Information. [Wikipedia]

JVM

Die Java Virtual Machine ist für Ausführung des Java-Bytecodes zuständig sowie für die Kommunikation mit der zugrunde liegenden Maschine und dem Betriebssystem. Da Java-Programme in einer JVM ausgeführt werden, sind sie weitestgehend Plattform-unabhängig.

Ontologie

Im Rahmen der KI-Forschung (KI - künstliche Intelligenz) Modellierung von Domänen der realen Welt mit dem Ziel eines strukturierten und fundierten Aufbaus von Wissensbasen, damit rechnergestützt maximales Wissen generiert werden kann.

Polyhierarchie

(Multihierarchie) bezeichnet im Zusammenhang mit Begriffssystemen wie Ontologien und Thesauri (Begriffshierarchien) eine hierarchische Struktur, in der eine Klasse mehr als eine übergeordnete Klasse haben kann. Somit kann jeder Begriff mehreren Oberbegriffen zugeordnet werden, da verschiedene Merkmale bei der Zuordnung berücksichtigt werden.

Prozess

Der Begriff des Prozesses wird wie folgt definiert: Eine Gesamtheit von aufeinander einwirkenden Vorgängen in einem System, durch die Materie, Energie oder Information umgeformt, transportiert oder gespeichert wird.³⁶

Regulärer Ausdruck

Eine Zeichenkette, die mit Hilfe einer festgelegten Syntax ein Muster formal beschreibt (z. B. das Format von deutschen Postleitzahlen oder gültigen Email-Adressen). Reguläre Ausdrücke werden in der Informatik häufig zur Suche von bestimmten Mustern in Texten oder zur Plausibilitätsprüfung von Benutzereingaben eingesetzt.

Schlagwortliste

Schlagworte werden aus dem Text heraus vergeben, teilweise von Experten und Laien, daher kann die Relevanz und Zuordnung unterschiedlich ausgelegt sein.

Semantik

Die **Semantik** (Bedeutungslehre) ist das Teilgebiet der Sprachwissenschaft (Linguistik), das sich mit **Sinn** und **Bedeutung** von Sprache beziehungsweise sprachlichen Zeichen befasst. Die Semantik kümmert sich um die Frage, wie Sinn und Bedeutung von komplexen Begriffen aus denen von einfachen Begriffen abgeleitet werden können und stützt sich dabei in der Regel auf die Syntax. Hierbei bezeichnet nach Gottlob Frege

- Sinn (engl. sense) den Inhalt, der sich aus den Relationen der Zeichen, Wörter, Sätze etc. untereinander im System der Sprache ergibt,
- Bedeutung (engl. reference) den Inhalt, der sich aus der Relation zwischen Zeichen und Welt ergibt.

Semantisches Netz

Ein semantisches Netz ist ein formales Modell von Begriffen und ihren Beziehungen (Relationen). Es wird in der Informatik und künstlichen Intelligenz zur Wissensrepräsentation genutzt.

SQL

Structured Query Language: Eine Datenbanksprache, die sich bei relationalen Datenbanken allgemein als gängiger Standard etabliert hat. Sie bietet Operationen zur Abfrage, zur Definition und zur Manipulation von Daten.

³⁶ K. Schmid, Komplexe interaktive Systeme in der anwendungsorientierten Informatik Interdisziplinäre Analyse-, Entwicklungs- und Gestaltungsmethoden, Stuttgart 1994, S. 9

Syntax

behandelt die Muster und Regeln, nach denen Wörter zu größeren funktionellen Einheiten wie Phrasen (*Teilsatz*) und Sätzen zusammengestellt und Beziehungen wie Teil-Ganzes, Abhängigkeit etc. zwischen diesen formuliert werden (*Satzbau*).

Taxonomie

Ordnung von Dingen, zumeist in Form einer hierarchischen Klassifikation.

Thesaurus

Ein Thesaurus bzw. Wortnetz ist in der Dokumentationswissenschaft ein kontrolliertes Vokabular, dessen Begriffe durch Relationen miteinander verbunden sind. Die Bezeichnung wird gelegentlich auch für linguistische Thesauri oder wissenschaftliche Wortschatz-Sammlungen verwendet. Der Ausdruck „Thesaurus“ stammt aus dem Griechischen (thesauros für Schatz, Schatzhaus) bzw. aus dem Lateinischen.

Wissensmanagement

Wissensmanagement beschäftigt sich mit den Möglichkeiten der Einflussnahme auf die Wissensbasis eines Unternehmens. Unter der Wissensbasis eines Unternehmens werden alle Daten und Informationen, alles Wissen und alle Fähigkeiten verstanden, die diese Organisation zur Lösung ihrer vielfältigen Aufgaben benötigt. Dabei werden individuelles Wissen und Fähigkeiten (Humankapital) systematisch in der Organisation verankert. Wissensmanagement kann daher als ein Interventionsmechanismus verstanden werden, der auf den Theorien der Organisationslehre und des OL (organisationales Lernen) beruht und diese systematisch nutzt.

Das Wissen innerhalb eines Unternehmens wird dabei als Produktionsfaktor verstanden, der neben Kapital, Arbeit und Boden tritt. Die strategische Grundlage für das Wissensmanagement bietet vor allem der Knowledge-based View of the Firm. Dieser stellt eine Erweiterung der Auffassung dar, Information (z. B. im Rahmen der Marktgestaltung und -beeinflussung) als betriebliche Ressource bzw. als Produktionsfaktor zu sehen.³⁷

³⁷ Quelle, Wikipedia, Abruf vom 29.05.2006, www.wikipedia.de

14 Literaturverzeichnis

14.1 Grundlagen

14.1.1 Mathematik

- [Diest06] Diestel, Reinhard: Graphentheorie. 3., neu bearb. und erw. Aufl. Springer, Berlin Heidelberg New York 2006. XVI, 344 S. ISBN 3-540-21391-0. <http://www.math.uni-hamburg.de/home/diestel/books/graphentheorie/GraphentheorieIII.pdf>
- [Lenze97] Lenze, Burkhard: Einführung in die Mathematik neuronaler Netze: mit C-Anwendungsprogrammen im Internet. Logos-Verlag, Berlin 1997. X, 249 S. ISBN 3-89722-021-0.

14.1.2 Begriffssysteme

- [DubCore] <http://dublincore.org/documents/1998/09/dces/>
- [MetaGd] <http://www2.sub.uni-goettingen.de/index.html>
- [SHMeta] <http://de.selfhtml.org/html/kopfdaten/meta.htm>
- [Voss04] Voß, Jakob: Begriffssysteme. Ein Vergleich verschiedener Arten von Begriffssystemen und Entwurf des integrierenden Thema-Datenmodells (Studienarbeit) . Humboldt Universität zu Berlin 2003. <http://www.jakobvoss.de/epub/begriffssysteme03/begriffssysteme.pdf>

14.2 Informatik

14.2.1 Informationssysteme

- [AltaV] AltaVista Overture Services Inc., Pasadena, California: <http://www.altavista.com/>
- [AskJee] IAC Search & Media Oakland, CA, Ask Jeeves: <http://www.ask.com/>
- [Brock] Bibliographisches Institut & F. A. Brockhaus AG, Mannheim: <http://www.brockhaus.de/>
- [Ferb05] Ferber, Reginald: Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. 1. Aufl. dpunkt, Heidelberg 2003. XI, 340 S. ISBN 3-89864-213-5. <http://information-retrieval.de/irb/ir.html>
- [Google] Google Inc. Mountain View, CA, USA: <http://www.google.de/>
- [Leo] Leo (Link Everything Online): <http://www.leo.org/>
- [Lewa05] Lewandowski, Dirk: Web Information Retrieval: Technologien zur Informationssuche im Internet (Reihe Informationswissenschaft der DGI; Bd. 7) [Hrsg.: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V., DGI] DGI, Frankfurt am Main 2005. 248 S. ISBN 3-925474-55-2. <http://www.durchdenken.de/lewandowski/web-ir/>
- [Media] mediaTUM: <http://mediatum.ub.tum.de/>
- [MSN] Microsoft Network, MSN: <http://www.msn.com/>
- [Panyr86] Panyr, Jiri: Automatische Klassifikation und Information retrieval: Anwendung u. Entwicklung komplexer Verfahren in Information-retrieval-Systemen u. ihre Evaluierung (Sprache und Information; Bd. 12). Niemeyer, Tübingen 1986. XII, 416 S. ISBN 3-484-31912-7.
- [Wikipedia] Wikipedia: <http://www.wikipedia.de/>
- [WitMül] Witte, René [Hrsg.]; Mülle, Jutta [Hrsg.]: Text Mining: Wissensgewinnung aus natürlichsprachigen Dokumenten. Universität Karlsruhe (TH), Fakultät für Informatik 2006.
- [Yahoo] Yahoo! Inc.: <http://www.yahoo.com/>

14.2.2 Programmierung

- [Ant] Apache Ant: <http://ant.apache.org/>
- [Cavan06] Cavaness, Chuck: Quartz job scheduling framework: building open source enterprise applications. Prentice Hall, Harlow 2006. xxii, 330 p. ISBN: 0-13-188670-3.
- [Eclipse] Eclipse Foundation Inc., Ottawa, Ontario, Canada: <http://www.eclipse.org/>
- [Fish03] Fisher, Maydene; Ellis, Jon; Bruce, Jonathan: JDBC API Tutorial and Reference. 3. ed. Addison Wesley 2003. 1242 S. ISBN 0-3211-7384-8.
- [Geary99] Geary, David M.: Graphic Java 2.0: Die JFC beherrschen (AWT). Bd. 1., 3. Aufl. Prentice Hall, München 1999. 926 S. ISBN 3-8272-9585-8.
- [Hall04] Hall, Marty; Brown, Larry: Core Servlets und JavaServer Pages: [Schritt-für-Schritt-Einstieg in die Entwicklung webfähiger Applikationen und dynamischer Webseiten; tiefgehende Besprechung der aktuellen Spezifikationen Servlets 2.4 und JSP 2.0 und der neuen Features; Profi-Details für die Arbeit mit Apache Tomcat, Macromedia JRun und Caucho Resin] [Übers.: Petra Alm]. 2. Ausg. Markt und Technik, München/Germany 2004. 624 S. ISBN 3-8272-6645-9.
- [Krue06] Krüger, Guido: Handbuch der Java-Programmierung. 4., aktualisierte. Aufl. Addison-Wesley München, Boston [u. a.] 2006. 1274 S. ISBN 3-8273-2361-4.
- [Musc03] Musciano, Chuck; Kennedy, Bill: HTML & XHTML: das umfassende Referenzwerk. Dt. Übers. von Eva Wolfram & Imke Schenk. 4., Aufl., dt. Ausg. der 5. Aufl. O'Reilly, Beijing Cambridge Farnham Köln Paris Sebastopol Taipei Tokyo 2003. XIX, 698 S. ISBN 3-89721-350-8.
- [Quartz] Job Scheduler (Open Source): <http://www.opensymphony.com/quartz/>
- [Ulle06] Ullenboom, Christian: Java ist auch eine Insel: das umfassende Handbuch; [Programmieren mit der Java Standard Edition Version 5]. 5. aktualisierte und erw. Auflage. Galileo Press, Bonn 2006. 1456 S. ISBN 3-89842-747-1.
- [Subver] Subversion: <http://subversion.tigris.org/>

14.2.3 Software Engineering

- [Alhir98] Alhir, Sinan Si: UML in a nutshell: a desktop quick reference. 1. ed. O'Reilly, Cambridge Köln Paris Sebastopol Tokyo 1998. XIII, 273 S. ISBN 1-56592-448-7.
- [Berli] BerliOS: Better SCM Initiative (Vergleich von Versionsverwaltungen): <http://better-scm.berlios.de/comparison/comparison.html>
- [Ernst03] Ernst, Hartmut: Grundkurs Informatik: Grundlagen und Konzepte für die erfolgreiche IT-Praxis; eine umfassende, praxisorientierte Einführung. 3., überarb. und verb. Aufl. Vieweg, Braunschweig Wiesbaden 2003. XIX, 888 S. ISBN 3-528-25717-2.
- [Free06] Freeman, Eric; Freeman, Elisabeth: Entwurfsmuster von Kopf bis Fuß: [ein Buch zum Mitmachen und Verstehen] Mit Kathy Sierra ; Bert Bates. Dt. Übers. von Lars Schulten und Elke Buchholz. 1. Aufl. O'Reilly, Beijing Cambridge Köln Paris Sebastopol Taipei Tokyo 2006. XXXIV, 638 S. ISBN 3-89721-421-0.
- [Gam04] Gamma, Erich; Riehle, Dirk [Übers.]: Entwurfsmuster: Elemente wiederverwendbarer objektorientierter Software. 1. Aufl., [Neuauf.]. Addison-Wesley, München Boston [u. a.] 2004. XX, 479 S. ISBN 3-8273-2199-9.
- [Oester] Oestereich, Bernd: UML-Notationsübersicht: <http://www.oose.de/notationuml14.htm>

14.2.4 Persistenz

- [HSQL] HSQL (in Java programmiert): <http://hsqldb.org/>
- [HSQL2] http://www.javamagazin.de/itr/online_artikel/psecom.id.312.nodeid.11.html
- [MySQL] MySQL: <http://www.mysql.de/>
- [Oracle] Oracle Corporation: <http://www.oracle.com/>
- [Petkov] Petkovič, Dusan: SQL: die Datenbanksprache. McGraw-Hill, Hamburg New York [u. a.] 1990. 249 S. ISBN 3-89028-178-8.
- [PoSQL] PostgreSQL: <http://www.postgresql.org/>

14.2.5 Netzwerk

- [RFC2246] RFC 2246: The TLS Protocol Version 1.0
- [RFC2396] RFC 2396: Uniform Resource Identifiers (URI): Generic Syntax
- [RFC2616] RFC 2616: HTTP
- [RFC2817] RFC 2817: Upgrading to TLS Within HTTP/1.1
- [RFC2818] RFC 2818: HTTP Over TLS
- [RFC3546] RFC 3546: Transport Layer Security (TLS) Extensions

14.2.6 Application Server

- [Gucker] Gucker, Joachim; Müller Michael [u. a.] (ARS Computer und Consulting GmbH): Webanwendungen mit IBM Rational und IBM WebSphere V6 : effizient entwickeln mit J2EE 1.4, JSF und SDO (IBM software press). Addison-Wesley, München Boston [u. a.] 2006. 944 S. ISBN 3-8273-2230-8.
- [JBoss] JBoss: <http://de.jboss.com/> <http://www.jboss.com/>
- [Tomcat] Apache Tomcat: <http://tomcat.apache.org/>
- [WebSph] IBM WebSphere: <http://www.ibm.com/websphere>

14.2.7 Indexierung

- [Berg01] Bergman, Michael K.: The Deep Web: surfacing hidden value. Journal of Electronic Publishing, 2001, Vol. 7, No. 1.
www.press.umich.edu/jep/07-01/bergman.html
- [Lewa06] Lewandowski, Dirk; Mayr, Philipp: Exploring the Academic Invisible Web. Library Hi Tech 24(2006)4.
http://www.durchdenken.de/lewandowski/doc/LHT_Preprint.pdf
- [Lucene] Lucene: <http://lucene.apache.org/>
- [MILOS] MILOS: http://www.ub.uni-duessel-dorf.de/home/ueber_uns/projekte/abgeschlossene_projekte/milos/index_html
- [Nohr03] Nohr, Holger: Grundlagen der automatischen Indexierung. Ein Lehrbuch. Logos-Verlag, Berlin 2003. 153 S. ISBN 3-8325-0121-5.
- [Nohr] Nohr, Holger: Automatische Dokumentenindexierung – Eine Basistechnologie für das Wissensmanagement. Reihe: Arbeitspapiere Wissensmanagement, Nr. 2/2000. Fachhochschule Stuttgart, 2000. 61 S. ISSN 1616-5349 (Internet), ISSN 1616-5330 (Print).
<http://www.iuk.hdm-stuttgart.de/nohr/KM/KmAP/Indexing.pdf>
- [Nutch] Nutch: <http://lucene.apache.org/nutch>
- [RedPi] Red Piranha: <http://red-piranha.sourceforge.net/>
- [Regain] Regain Hilfe: <http://regain.murfman.de/wiki/de/index.php/Hauptseite>
- [Schn1] Schneider, Tilman: Effizientes Suchen mit Jakarta Lucene (Regain):
http://regain.sourceforge.net/download/Seminararbeit_Lucene_Ausarbeitung.pdf
- [Schn2] Schneider, Tilman: Jakarta Lucene, Eine Java-Bibliothek zur Suchindex-Erstellung (Regain) Folien zur Seminararbeit:
http://regain.sourceforge.net/download/Seminararbeit_Lucene_Vortrag.pdf
- [Schn3] Schneider, Tilman: Anwenderhandbuch Regain:
http://mesh.dl.sourceforge.net/sourceforge/regain/Anwenderhandbuch_Regain_v1.0.pdf

14.2.8 Semantisches Netz

- [Anto04] Antoniou, Grigoris.; Van Harmelen, Frank: A Semantic Web Primer. MIT Press 2004. 272 p. ISBN 0262012103.
- [Beats] Beats Biblionetz: <http://beat.doebe.li/bibliothek>
- [Bern01] Berners-Lee, Tim; Hendler, James A.; Lassila, Ora: The Semantic Web. Scientific American 284(5):34-43, May 2001. http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21
- [Daco03] Daconta, Michael C.; Oberst, Leo J.; Smith, Kevin T.: The Semantic Web. 1. ed. John Wiley & Sons 2003. 312 p. ISBN 0-471-43257-1.
- [Gure05] Gurevych, Iryna: Anwendungen des semantischen Wissens über Konzepte im Information Retrieval. EML Research gGmbH, Heidelberg 2005. <http://www.eml-r.villa-bosch.de/english/homes/gurevych/downloads/knowledge05.pdf>
- [Jena] Jena, Framework für Java: <http://jena.sourceforge.net/>
- [JWebSeb] Journal of Web Semantics http://www.elsevier.com/wps/find/journaldescription.cws_home/671322/description#description
- [Maed02] Maedche, Alexander: Ontology Learning for the Semantic Web. 1. ed. Kluwer Academic Publishers 2002. 272 p. ISBN 0792376560.
- [OpThes] Open Thesaurus: <http://www.openthesaurus.de/>
- [OWL] OWL: <http://www.w3.org/TR/owl-features/>
- [RDF1] RDF: <http://www.w3.org/RDF>
- [RDFN3] Tim Berners-Lee: Getting into RDF & Semantic Web using N3: <http://www.w3.org/2000/10/swap/Primer.html>
- [RDFV] RDF-Validator: <http://www.w3.org/RDF/Validator>
- [SWSer] Semantik und Web Services: Vokabulare und Ontologien: <http://www.jeckle.de/semanticWebServices/vokont.html>
- [SWSW] Semantic Web School in Wien: <http://www.semantic-web.at/>
- [Tani06] Taniar, David (Hrsg.); Rahayu, Johanna Wenny (Hrsg.): Web Semantics and Ontology. Idea Group Publishing 2006. 404 p. ISBN 1-59140-906-3.
- [Tolks06] Tolksdorf, Robert; Paslaru, Elena. Tutorial: Grundlagen des Semantic Web: http://www.inf.fu-berlin.de/inst/ag-nbi/lehre/0506/P_SW/fohlen/rdfs2.pdf

14.2.9 Visualisierung

- [EXIF] EXIF: <http://www.exif.org/Exif2-2.PDF>
- [FreeM] FreeMind (Java): http://freemind.sourceforge.net/wiki/index.php/Main_Page
- [GraVi] GraphViz: <http://www.graphviz.org>
- [IsaVi] IsaViz: <http://www.w3.org/2001/11/IsaViz>
- [JGrap] JGraph: <http://www.jgraph.com/>
- [JView] JViews: <http://www.ilog.com/products/jviews>
- [kdiss] kdissert: <http://www.kde.org.uk/apps/kdissert/>
- [Kgrap] Kgraphviewer: <http://www.kde-apps.org/content/show.php?content=23999>
- [Monar] MonarchGraph: <http://www.singleton-labs.com/>
- [Prote] Protégé: <http://protege.stanford.edu/>
- [SunImgiO] Sun: ImageIO für Java: http://java.sun.com/products/java-media/jai/forDevelopers/jai-imageio-1_0-rc-docs/overview-summary.html
- [TomSa] Tom Sawyer Visualization, Java Edition: <http://www.tomsawyer.com/tsv/tsv.java.php>
- [TouchG] TouchGraph (dynamische Ausrichtung der Begriffe): <http://www.touchgraph.com>
- [VYM] VYM – View Your Mind: <http://www.kde-apps.org/content/show.php?content=26420>

14.2.10 Lernen

- [Duda01] Duda, Richard O.; Hart, Peter E.; Stork, David G.: Pattern classification. 2. ed. Wiley, New York 2001. xx, 654 p. ISBN 0471056693.
- [Ester00] Ester, Martin; Sander, Jörg: Knowledge discovery in databases: Techniken und Anwendungen. Springer, Berlin Heidelberg New York Barcelona Hongkong London Mailand Paris Singapur Tokio 2000. VIII, 281 S. ISBN 3-540-67328-8.
- [Hastie01] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome: The elements of statistical learning: data mining, inference, and prediction. Springer, New York 2001. xvi, 533 p. ISBN 0387952845.
- [Hein99] Heinsohn, Jochen ; Socher-Ambrosius, Rolf: Wissensverarbeitung: eine Einführung (HochschulTaschenbuch). Spektrum, Akad. Verl., Heidelberg Berlin 1999. XVI, 310 S. ISBN 3-8274-0308-1.
- [Heyer02] Heyer, G.; Quasthoff, U.; Wolff, Chr.: Möglichkeiten und Verfahren zur automatischen Gewinnung von Fachbegriffen aus Texten. Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig 2002. <http://wortschatz.informatik.uni-leipzig.de/asv/publikationen/HeyFachbegriffe100902.pdf>
- [IIF] Institut für Informatik Freiburg: Arbeitsgruppe Grundlagen der Künstlichen Intelligenz. Publikationen zu AI: http://www.informatik.uni-freiburg.de/~ki/publications/by_area_de.html
- [KernMach] Kernel-Machines.Org: <http://www.kernel-machines.org/>
- [Kloes02] Kloesgen, Willi; Zytkow, Jan: Handbook of data mining and knowledge discovery. Oxford University Press, Oxford 2002. xxii, 1026 p. ISBN 0195118316.
- [Maim05] Maimon, Oded Z.; Rokach, Lior: Data mining and knowledge discovery handbook. Springer, Ramat-Aviv ; [Great Britain] : 2005. xxxv, 1383 p. ISBN 0387244352.
- [Mitch97] Mitchell, Tom M.: Machine Learning. McGraw-Hill, New York 1997. xvii, 414 p. ISBN 0-07-042807-7
- [Russ98] Russell, Stuart J.; Norvig, Peter: Artificial intelligence: a modern approach. 2. ed. Prentice Hall, Upper Saddle River, N. J. 1998. xxviii, 1081 p. ISBN 0137903952.
- [Schoel02] Schölkopf, Bernhard; Smola, Alexander J.: Learning with kernels: support vector machines, regularization, optimization, and beyond (Adaptive computation and machine learning) MIT Press, Cambridge London 2002. xviii, 626 p. ISBN 0262194759. <http://www.learning-with-kernels.org/>
- [Shawe04] Shawe-Taylor, John; Cristianini, Nello: Kernel methods for pattern analysis. Cambridge : Cambridge University Press, 2004. xiv, 462 p. ISBN 0521813972. <http://www.kernel-methods.net/>
- [Weik04] Weikum, Gerhard: Effiziente Informationssuche im Web der Zukunft. Max-Planck-Institut für Informatik, Saarbrücken, Abteilung Datenbanken und Informationssysteme. <http://www.mpg.de/bilderBerichteDokumente/dokumentation/jahrbuch/2004/informatik/forschungsSchwerpunkt1/pdf.pdf>
- [Witten05] Witten, Ian H.; Frank, Eibe: Data mining: practical machine learning tools and techniques. 2nd ed. (Morgan Kaufmann series in data management systems). Morgan Kaufmann, San Francisco, Calif.; Elsevier Science [distributor], Oxford 2005. xxxi, 525 p. ISBN 0120884070.

14.2.11 Dokumentenmanagement

- [CnttMngr] Contentmanager.de - Das Content Management Portal:
<http://www.contentmanager.de/>
- [Cocoon] Apache-Projekt Cocoon (Transformation von Dokumenten):
<http://cocoon.apache.org/>
- [Hauss03] Haussner, Stephanie: Kooperativ an Webseiten arbeiten mit einem WikiWikiWeb. Pädagogische Hochschule Heidelberg 2003:
http://www.ph-heidelberg.de/org/mediatut/wiki_skript.pdf
- [Lenya] Apache-Projekt Lenya (Open Source Content Management, Java, XML): <http://lenya.apache.org/>
- [Type3] Typo3: <http://typo3.org/>
- [WCmp] Vergleich von Wiki-Engines: <http://www.wikimatrix.org/>
- [WCmp2] Vergleich von Wiki-Engines:
<http://moinmoin.wikiwikiweb.de/WikiEngineComparison>
- [WFrik] Frik (Java): <http://sourceforge.net/projects/friki>
- [Wlke] IkeWiki (Java): <http://sourceforge.net/projects/ikewiki/>
- [WJSP] JSPWiki (Java): <http://www.jspwiki.org/>
- [WLst] Auflistung von Wiki-Software: <http://de.wikipedia.org/wiki/Wiki>
- [WLst2] Auflistung von Wiki-Software: <http://www.c2.com/cgi/wiki?WikiEngines>
- [WMed] MediaWiki (Wikipedia und Co., PHP):
<http://de.wikipedia.org/wiki/MediaWiki>
- [WMoin] MoinMoin (Python): <http://sourceforge.net/projects/moin>
- [WSem] Semantic Wiki:
http://wiki.ontoworld.org/index.php/Semantic_Wiki_State_Of_The_Art
- [WSem2] Semantic Wiki: http://en.wikipedia.org/wiki/Semantic_Wiki
- [WSnip] SnipSnap (Java): <http://snipsnap.org/space/start>
- [WUseM] UseModWiki (Perl): <http://www.usemod.com/cgi-bin/wiki.pl>
- [WVQ] Very Quick Wiki (Java): <http://www.vqwiki.org>

14.2.12 Abrechnungssysteme, Affiliate, Foren und Weblogs

- [AdGgl] Google AdSense: <https://www.google.com/adsense>
- [AdSen] AdSense: <http://de.wikipedia.org/wiki/AdSense>
- [AdWrd] Google AdWords: <http://de.wikipedia.org/wiki/AdWords>
- [AffAc] Affiliate Academy: <http://www.affiliate-academy.de/>
- [AffAm] Amazon: <http://partnet.amazon.de/gp/associates/join/main.html>
- [AffDe] Affiliate.de: <http://www.affiliate.de/>
- [Affil] Affiliate: <http://de.wikipedia.org/wiki/Affiliate>
- [AffJat] Jatismo Partnerprogramme:
<http://www.jatismo.de/partnerprogramme.html>
- [AffSW] Affilisoft: <http://www.affilisoft.de/>
- [BlogS] Google Blogsearch: <http://blogsearch.google.com/>
- [Blogs] Weblogs, FAQ: <http://www.stefanbucher.net/weblogfaq>
- [CetB] Click&Buy: <http://clickandbuy.com/DE/de/>
- [Foren] Foren: <http://aktuell.de.selfhtml.org/artikel/gedanken/foren-boards>
- [ForSW] Forensoftware: <http://www.forensoftware.de/index.php>
- [ZahlSys] Universität Konstanz, Informationswissenschaften: Zahlungssysteme im Internet: <http://www.inf-wiss.uni-konstanz.de/CURR/summer98/imk/Internet-Zahlungssysteme/zahlungssysteme.html>

14.3 Holzbau

14.3.1 Internet-Präsenzen

- [Dataholz] Dataholz.com: <http://www.dataholz.com/>
- [DBHolz] Datenbank Holz des ihd Dresden: <http://www.dbholz.de/>
- [DGfH] DGfH (Deutsche Gesellschaft für Holzforschung): <http://www.dgfh.de/>
- [HolzDE] Holz.de: <http://holz.de/>
- [HolzDB] Datenbank Holzeigenschaften: <http://www.holzdatenbank.de/>
- [IFT] Institut für Fenstertechnik e. V.: <http://www.ift-rosenheim.de/>
- [Infoholz] Infoholz.de: <http://www.infoholz.de/>
- [InstArch] Institut für Architekturwissenschaften: <http://www.iti.tuwien.ac.at/>
- [IRB] IRB-Verlag: <http://www.irbdirekt.de/>
- [WoodPort] A Portal for Forestry, Wood and Furniture Research and Training:
<http://www.innovawood.com/>
- [ZmrMstr] Zimmermeister: <http://www.zimmermeister-suche.de/>

14.3.2 Patentwesen

- [DPMA] Deutsches Patent- und Markenamt: <http://www.dpma.de/>
- [EuroPat] Patentwesen: <http://www.epo.org/>

14.3.3 Statistik

- [IuKBW05] R. Kössler, „Informations- und Kommunikationstechnologie in Unternehmen 2005“, Statistisches Monatsheft Baden-Württemberg 3/2006
- [SBaPK06] Statistisches Bundesamt, Pressemitteilung, „Aktuelle Nutzung von Informationstechnologie in Deutschland“ am 21.02.06 Frankfurt Main

14.3.4 Sonstiges

- [SK94] K. Schmid, Komplexe interaktive Systeme in der anwendungsorientierten Informatik Interdisziplinäre Analyse-, Entwicklungs- und Gestaltungsmethoden, Stuttgart 1994, S. 18
- [SM03] M. Seidel, Die Bereitschaft zur Wissensteilung, Rahmenbedingungen für ein wissensorientiertes Management, 1. Auflage, Deutscher Universitäts-Verlag, Wiesbaden 2003

15 Anhang

15.1 Technische Systeminformationen

15.1.1 Gesamtsystem

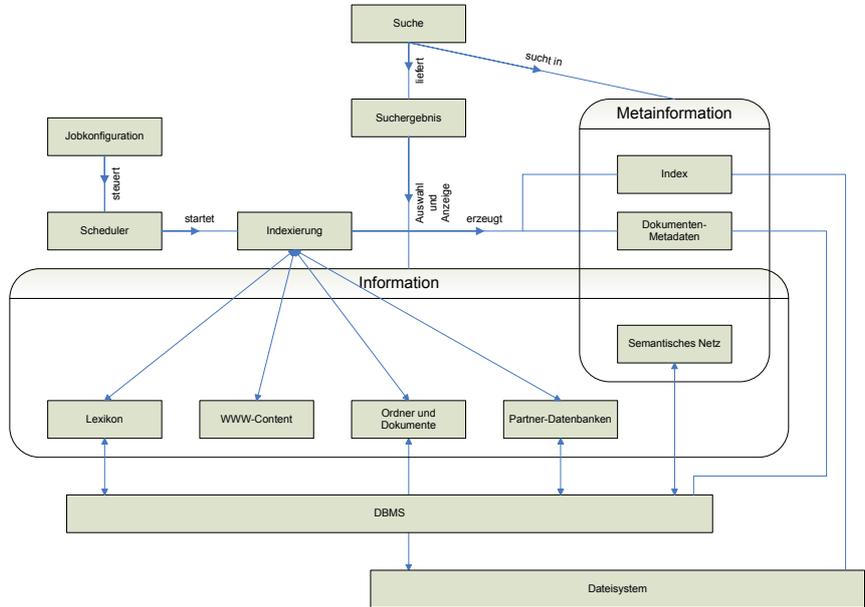


Abbildung 68: Überblick Gesamtsystem

15.1.2 Unabhängige Teilanwendungen

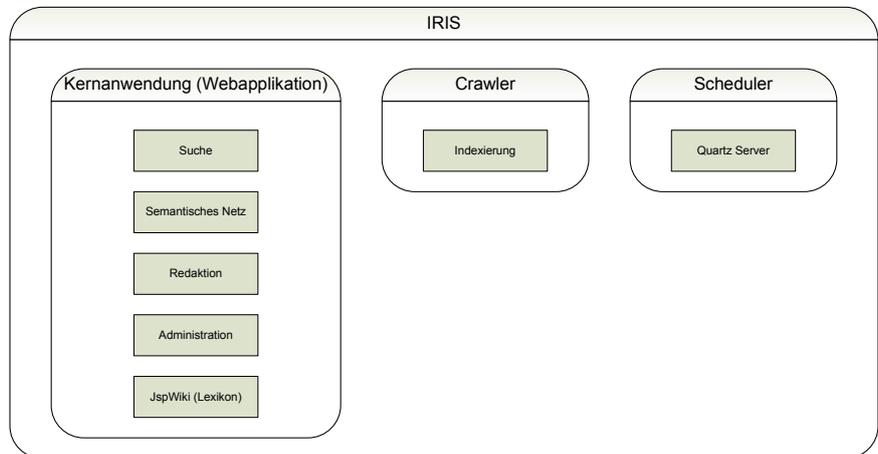


Abbildung 69: IRIS Teilanwendungen

Das Gesamtsystem IRIS besteht aus den drei unabhängigen Teilanwendungen Webapplikation, Crawler und Scheduler. Als eigenständige Programme laufen sie getrennt voneinander jeweils in einer eigenen Instanz der Java Virtual Machine (JVM).

15.1.3 Suche

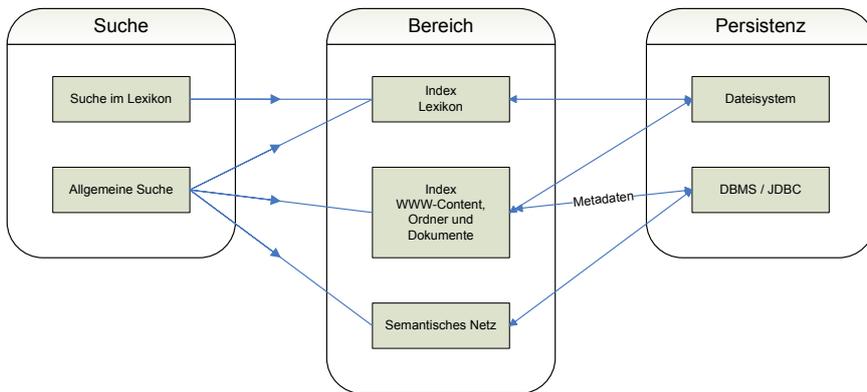


Abbildung 70: Zugriffe bei der Suche

15.1.4 Kommunikation zwischen Kernsystem und Lexikon

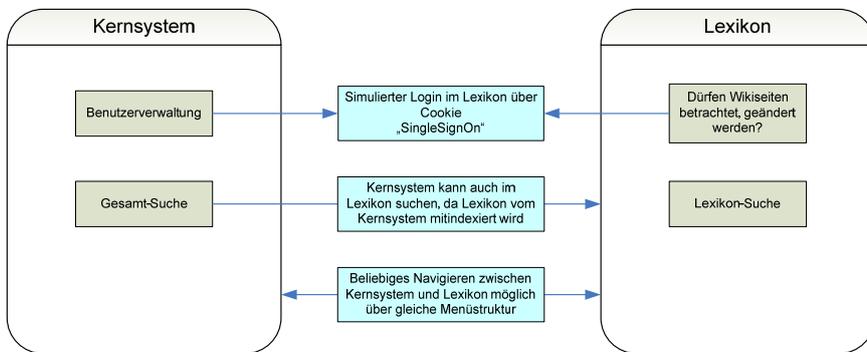


Abbildung 71: Kommunikation zwischen Kernsystem und Lexikon

15.1.5 Technologien

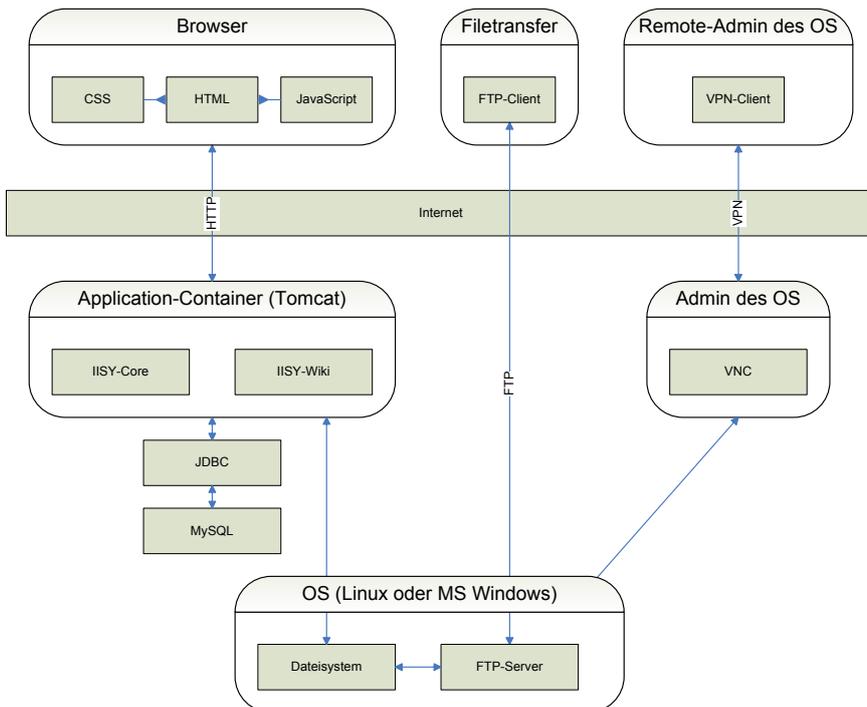


Abbildung 72: Eingesetzte Technologien

15.1.6 Ordner und Dokumente

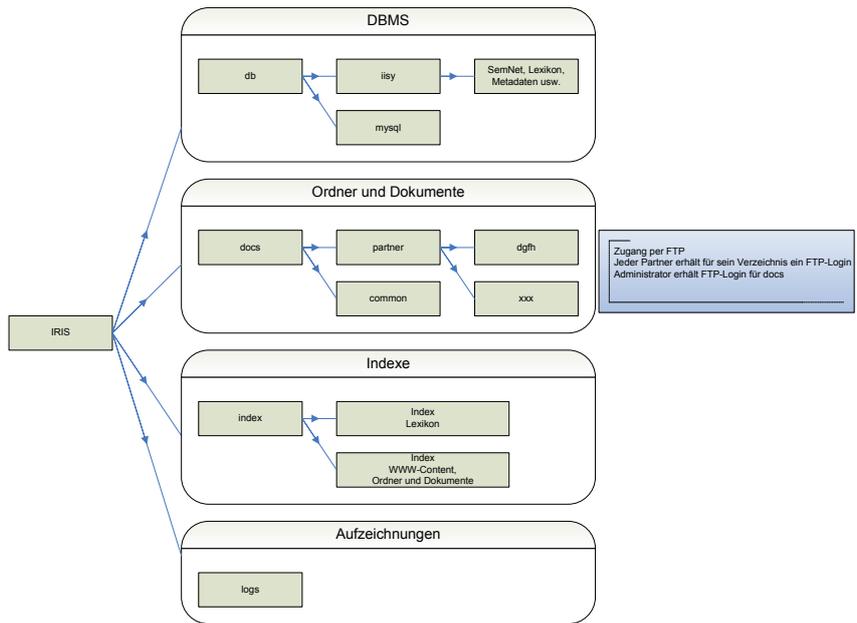


Abbildung 73: Struktur von Ordnern und Dokumenten

15.1.7 Umsetzung des MVC-Paradigmas (Model-View-Controller)

MVC in IRIS

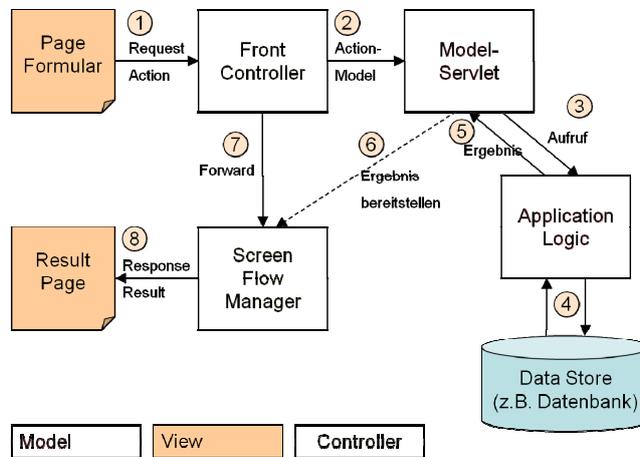


Abbildung 74: Verarbeitung einer Http-Anfrage in IRIS

Eine Anfrage (Schritt 1) wird vom Front Controller entgegen genommen. Dieser ruft gemäß der gewünschten Aktion das zuständige Model-Servlet auf (2). Dieses nimmt die Formulardaten entgegen und ruft damit die eigentliche Anwendungslogik auf (3). Die Anwendungslogik kann bei Bedarf auf die persistente Datenhaltung, z. B. eine Datenbank, zurückgreifen (4) und gibt das Ergebnis an das aufrufende Model-Servlet zurück (5). Dieses hinterlegt das Ergebnis in geeigneter Form für die spätere Präsentation (6). Vereinzelt kommt für komplexe Anfragen auch der Aufruf mehrerer Model-Servlets nacheinander in Betracht. Nach Erledigung dieser Aufgaben gibt der Front Controller die Kontrolle an den Screen-Flow Manager ab (7). Dieser ermittelt anhand der Anfrage und des zugehörigen Ergebnisses, welche Antwort (z. B. Ergebnisauflistung oder Fehlerseite) dem Anwender präsentiert werden soll (8). In den Ablauf sind Plausibilitätsprüfungen von Nutzereingaben sowie die

Kontrolle von notwendigen Benutzerrechten und entsprechende Reaktionen integriert.

15.2 Anwendungsfälle

15.2.1 Akteure und Anwendungsfälle im Überblick

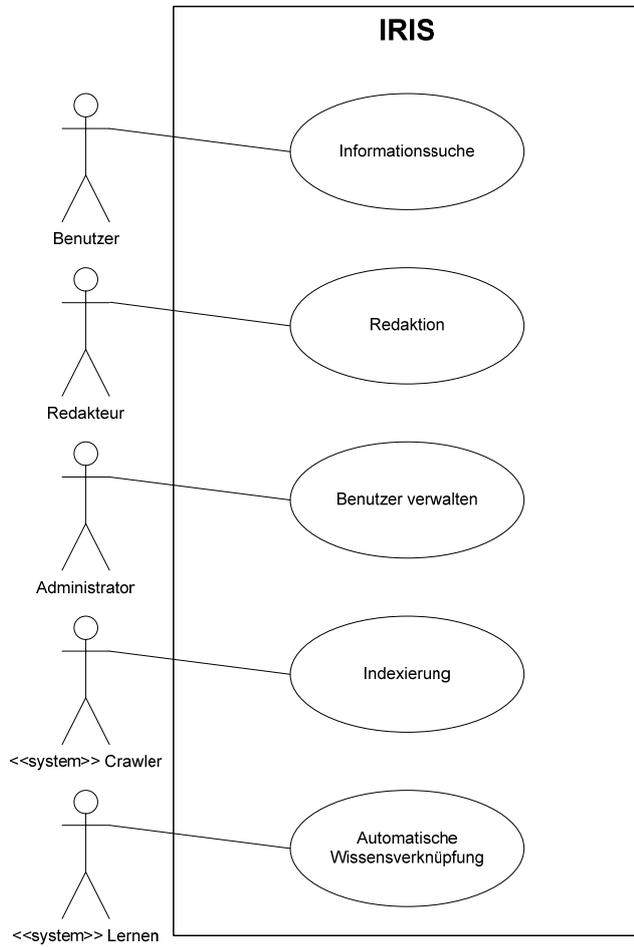


Abbildung 75: Akteure und Anwendungsfälle im Überblick

15.2.2 Anwendungsfälle der Benutzer

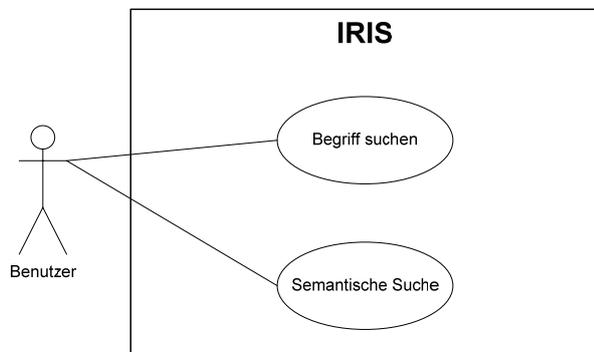


Abbildung 76: Anwendungsfälle der Benutzer

15.2.3 Anwendungsfälle der Redakteure

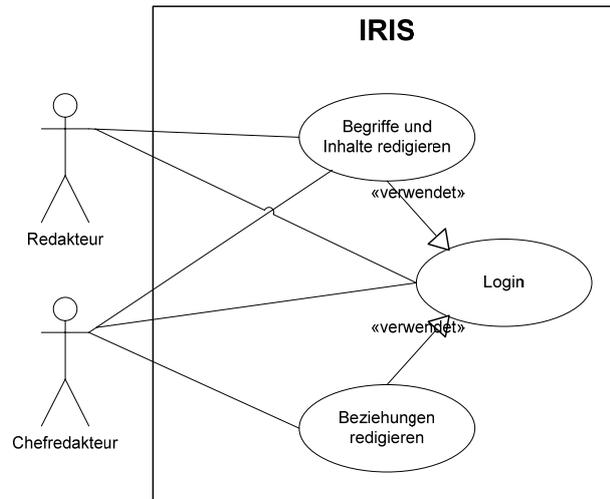


Abbildung 77: Anwendungsfälle der Redakteure

15.2.4 Anwendungsfälle des Administrators

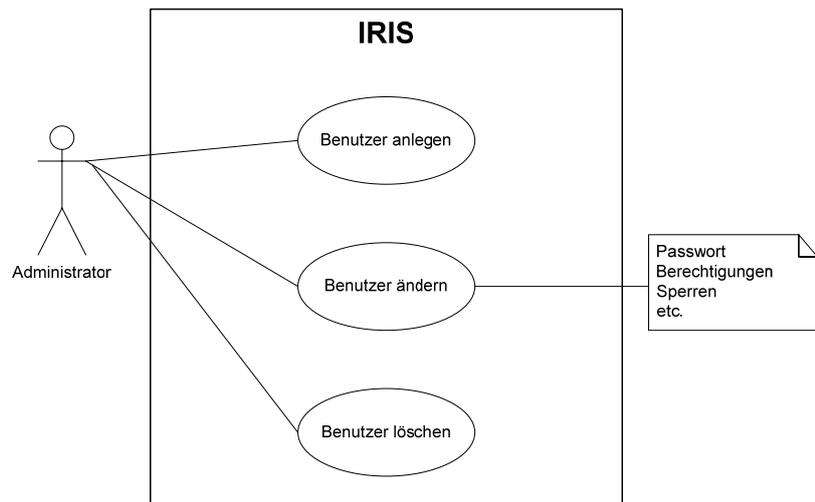


Abbildung 78: Anwendungsfälle des Administrators

	Datenbeständen
iisy.data.db	Implementierung von iisy.data mit einer Datenbank als Datenspeicher; Hilfsklassen
iisy.data.impl	Abstrakte Implementierung der Basisklasse aller persistent zu speichernden Objekte
iisy.err	Systemweit benötigte Ausnahmen
iisy.graphics	Einzelne grafische-Objekte des semantischen Netzes
iisy.graphics.test	Test für iisy.graphics
iisy.model.dbs	Operationen und Filter für integrierte Fremddatenbank
iisy.model.dbs.impl	Implementierungen von iisy.model.dbs
iisy.model.docs	Operationen und Filter für formatfreie Dokumenten-Abbildungen in der Datenbank
iisy.model.docs.impl	Implementierungen von iisy.model.docs
iisy.model.partner	Operationen und Filter für Projektpartner und Mitgliedschaften
iisy.model.partner.impl	Implementierungen von iisy.model.partner
iisy.model.person	Operationen und Filter für Benutzer
iisy.model.person.impl	Implementierungen von iisy.model.person
iisy.model.regainsetup	Operationen und Filter für verwaltete Indizierungskonfigurationen
iisy.model.regainsetup.impl	Implementierungen von iisy.model.regainsetup
iisy.model.semnet	Operationen und Filter für das semantische Netz (Begriffe, Prädikate, Beziehungen)
iisy.model.semnet.builder	Interface einer Klasse zur konkreten Erzeugung eines Ausschnittes des semantischen Netzes als Modell, textliche und grafische Darstellung
iisy.model.semnet.builder.impl	Implementierung von iisy.model.semnet.builder
iisy.model.semnet.impl	Implementierungen von iisy.model.semnet
iisy.model.semnet.learn	Operationen und Hilfsklassen für den Lernvorgang im semantischen Netz
iisy.model.semnet.learn.impl	Implementierungen von iisy.model.semnet.learn
iisy.model.semnet.search	Operationen und Hilfsklassen für die Termuche im semantischen Netz
iisy.model.semnet.search.impl	Implementierung von iisy.model.semnet.search
iisy.model.semnet.test	Tests für einzelne Funktionen im semantischen Netz
iisy.model.semnet.visual	Erzeugung einer grafischen Darstellung eines semantischen Netzes
iisy.model.semnet.visual.impl	Implementierungen von iisy.model.semnet.visual
iisy.model.tasks	Operationen und Filter für Automatisierte Aufgaben (Quartz Jobs); Hilfsklassen
iisy.model.tasks.impl	Implementierungen von iisy.model.tasks
iisy.model.wordlist	Operationen und Filter für Wortlisten
iisy.model.wordlist.impl	Implementierungen von iisy.model.wordlist
iisy.mvc	Hilfsklassen zur Umsetzung des MVC-Paradigmas in einer Webanwendungen: Front Controller, Screen Flow Manager, Dispatcher-Klassen, CoreEngine zur Integration der Dispatcher, Operations- und

	Kommunikationskonstanten, Filter auf Datenbestände, weitere Hilfsklassen
iisy.mvc.basicOps	Abstrakte Implementierungen der in der Anwendung vorgesehenen Standardoperationen auf persistente Datenbestände
iisy.mvc.filters	Web-Filter der Anwendung (nicht zu verwechseln mit den sonstigen logischen Filtern auf Datenbestände)
iisy.mvc.impl	Standardimplementierung eines Filters auf Datenbestände
iisy.opts	Operationen für Benutzereinstellungen und -optionen
iisy.opts.impl	Implementierung von iisy.opts
iisy.rsrc	Ressourcen-Dateien und Klasse Rsrc zur Einbindung dieser in das System
iisy.sched	Quartz Scheduler Client
iisy.sched.jobs	Eigene Quartz-Job Implementierungen
iisy.sched.server	Quartz Scheduler Server und Engine für die Abwicklung der Konfiguration
iisy.search	Operationen und Filter für die generelle Suchfunktion des Systems; Hilfsklassen zur Überleitung von Regain-Treffern in IRIS-Dokumenten-Einträge und Zugriffskontrolle
iisy.search.impl	Implementierung eines Suchfilters
iisy.settings	Benutzereinstellungen des semantischen Netzes
iisy.settings.impl	Implementierung von iisy.settings
iisy.util	Sammlung allgemein nützlicher Hilfsmittel und Funktionen: Zahlen- und Textkonvertierungen, einfacheres Handling von Arrays, Cookies, Dateien und Encodierungen, Versenden von Emails u. a.
iisy2regain	Hilfsklassen zur Anbindung der Regain-Suche von IRIS aus

15.4.2 Projektstruktur

Das Gesamtsystem IRIS ist in mehrere Teilprojekte unterteilt, die mit der Entwicklungsumgebung Eclipse erstellt wurden. Sie enthalten in der Summe den gesamten Quellcode sowie alle weiteren benötigten Bibliotheken und sonstigen Dateien sowie die Dokumentation. Alle Projekte sind auf dem Server mit der freien Dateiversionsverwaltungssoftware Subversion hinterlegt.

Tabelle 15: Quellcode Teilprojekte

Teilprojekt	Kurzbeschreibung
iisy-analyzers	IRIS-spezifische Analysierer, Wortlisten
iisy-base	Basis-Klassen der Anwendung: <ul style="list-style-type: none"> • Basisfunktionen zur Umsetzung des MVC-Pattern und Datenbankanbindung • Standardisierung von Modell-Operationen durch abstrakte Basisklassen • Textressourcen • Allgemeine Utilities (Helferklassen)
iisy-core	Kern der Webanwendung: <ul style="list-style-type: none"> • Ticket-, Partner- und Personen-Verwaltung • Nutzereinstellungen • Suchfunktion (Verbindung zu Regain)

	<ul style="list-style-type: none"> • Weboberfläche (JSP Dateien, HTML-Dateien, Bild- und Grafikdateien, web.xml) • SQL- und Shell-Scripte
lisy-documentation	Kein Quellcode, sondern Unterlagen zur Projekt- und System-Dokumentation
lisy-documents	Formatfreie Abbildung von Dokumenten in der Datenbank
lisy-libs	Die verwendeten Java-Bibliotheken einschließlich der IRIS-eigenen Jar-Dateien
lisy-regain	Für IRIS angepasste Version von Regain: <ul style="list-style-type: none"> • Angepasste build.xml zur Erzeugung v.a. des IRIS-Crawler • Shell-Script bzw. Batch-Datei für IRIS-Crawler • Konfigurationsvorlagen in XML • Mehrere fertige Konfigurationsdateien • Details zu Quellcode-Modifikationen siehe unten
lisy-regainsetup	Verwaltung von Crawler-Konfigurationsdateien
lisy-semnet	Das semantische Netz
lisy-tasks	Verwaltung automatisierter Aufgaben, <ul style="list-style-type: none"> • Client • Server • SQL-Script zur Erzeugung der nötigen Tabellen für persistente Haltung der Aufgaben
lisy-tools	Kein eigentlicher Teil von IRIS, sondern Sammlung kleiner nützlicher Werkzeuge, u. a.: <ul style="list-style-type: none"> • Test einzelner Teilmodule • WebClient • Xml2Html: transformiert XML- in HTML-Dateien mit farblicher Absetzung (für Konfigurationsvorlagen in IRIS aus den XML-Vorlagen im Projekt lisy-regain) Build-Datei zur Übersetzung und Erzeugung des gesamten Systems mit allen Teilanwendungen
jspwiki_iisy_jdbcprovider	Für IRIS fehlerbereinigte JDBC-Datenbank-anbindung für das WIKI-System
jspwiki_iisy_patches	Für IRIS angepasste Version von JSPWiki zur Realisierung der Enzyklopädie: <ul style="list-style-type: none"> • Patches zur Anpassung • Tomcat Konfigurationsdateien • Originalquellcode von JSPWiki Beim Patch-Vorgang (per Ant via beigefügter Build-Datei) wird der Originalquellcode in ein neues Projekt iisy_wiki kopiert und dann die vorhandenen Patches hinzugefügt bzw. gleichnamige Klassen mit den eigenen Versionen ersetzt. Das Resultat ist das IRIS-spezifische JSPWiki.

Die nachfolgende Übersicht zeigt die Verwendung der einzelnen Projekte in den Teilanwendungen.

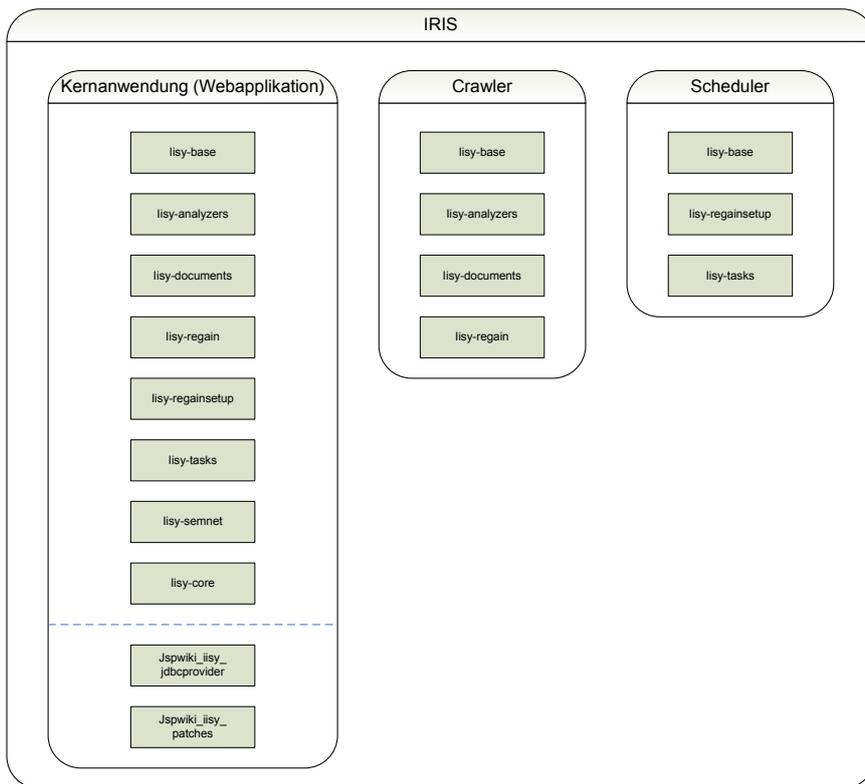


Abbildung 80: Zuordnung von Projekten zu Teilanwendungen

In der Webanwendung werden vom Projekt iisy-regain die Suchfunktionen auf dem Index und vom Crawler die Funktionen für das Crawling benötigt. Daher werden beim Projekt iisy-regain dementsprechend beim Übersetzen zwei Bibliotheken erzeugt: regain-search.jar und regain-crawler.jar. Das Modul JSPWiki ist relativ eigenständig und deswegen in Abbildung 80 mit einer gestrichelten Linie von den restlichen Projekten separiert. Bezüglich der Projektabhängigkeiten und der daraus resultierenden Implikationen für den Übersetzungsprozess siehe Anhang 15.5.

15.4.3 Modifizierte Software von Drittanbietern

Tabelle 16: Modifizierte Fremdsoftware

Fremdprojekt	Version	Kurzbeschreibung
Jdbcprovider	2007-03-05	Funktionalität zur Haltung von Seiten aus JSPWiki in einer Datenbank via JDBC
JSPWiki	2.4.53	Auf Java Server Pages (JSP) basierendes WIKI-System
Regain	1.2.3	Einfache Suchmaschine und Crawler in Java

15.4.4 Bibliotheken von Drittanbietern

Tabelle 17: Verwendete Software-Bibliotheken

Bibliothek (Dateiname)	Version	Anbieter	Kurzbeschreibung
activation.jar	1.1	Sun Microsystems, Inc.	JavaBeans(TM) Activation Framework
commons-codec-1.3.jar	1.3	Apache Software Foundation	Jakarta Commons Codec
commons-collections-3.1.jar	3.1	Apache Software Foundation	Jakarta Commons Collections

commons-lang-2.3.jar	2.3	Apache Software Foundation	Jakarta Commons Lang
commons-logging-1.1.jar	1.1	Apache Software Foundation	Jakarta Commons Logging
commons-pool-1.2.jar	1.2	Apache Software Foundation	Jakarta Commons Pool
j2ee.jar	1.4	Sun Microsystems, Inc.	Java 2 Platform, Enterprise Edition
jakarta-regexp-1.4.jar	1.4	Apache Software Foundation	Jakarta Regexp
junit-4.1.jar	4.1	http://junit.org (kein direkter Anbieter)	JUnit Framework
log4j-1.2.13.jar	1.2.13	Apache Software Foundation	Log4j zum Logging
lucene-analyzers-2.2.0.jar	2.2.0	Apache Software Foundation	Lucene Analysierer
lucene-core-2.2.0.jar	2.2.0	Apache Software Foundation	Lucene Kernapplikation
mail.jar	1.4	Sun Microsystems, Inc.	JavaMail(TM) API
mysql-connector-java-3.1.12-bin.jar	3.1.12	MySQL AB	MySQL Datenbankbindung via JDBC
quartz-all-1.5.2.jar	1.5.2	OpenSymphony	Quartz Framework zur Arbeitsablaufplanung
xercesImpl.jar	2.7.1	Apache Software Foundation	Xerces-J zum Parsen/Generieren von XML-Dateien

15.4.5 Quellcode-Anpassungen der verwendeten Fremdprojekte

Die nachfolgenden Auflistungen zeigen die im Vergleich zum Original vorgenommenen Veränderungen im Quellcode der verwendeten Fremdsoftware. Es wird nach Farben unterschieden dargestellt, welche Teile in der für IRIS entstandenen Version gänzlich neu hinzugekommen, welche in modifizierter Form und welche unverändert übernommen worden sind.

Tabelle 18: Farbcodierung zur Darstellung von Quellcode-Modifikationen

Farbcode	Erklärung
Neu	Die Klasse oder das Paket existiert im originalen Quellcode nicht, ist also für IRIS neu hinzugekommen.
Modifiziert	Im Quellcode der Klasse wurden Modifikationen (Änderungen oder Hinzufügungen) vorgenommen bzw. mindestens eine Klasse im Paket wurde modifiziert.
Unverändert	Die Klasse bzw. das Paket wurden nicht verändert.
Entfernt	Die Klasse wurde komplett entfernt. Hierzu gibt es einen gesonderten Kommentar.

15.4.5.1 Modifikationen in JdbcProvider (jspwiki_iisy_jdbcprovider)

Tabelle 19: Modifikationen am Quellcode von JdbcProvider

Paket	Klasse
com.forthgo.jspwiki.jdbcprovider	
	C3POConnectionProvider
	ConnectionProvider
	DBCPCConnectionProvider

	JDBCAttachmentProvider
	JDBCBaseProvider
	JDBCConnectionProvider
	JDBCPageProvider
	JDBCProviderConfiguration
	JNDIConnectionProvider
	PoolingJDBCConnectionProvider

Die beiden gänzlich entfernten Klassen C3POConnectionProvider und DBCPConnectionProvider importieren fremde Java-Klassen aus Drittbibliotheken, die in keinem der anderen IRIS Teilprojekte verwendet werden. Sie stellen lediglich zwei weitere Implementierungen von Protokollen (C3PO bzw. DBCP) neben JDBC dar, die im IRIS-Umfeld ohnehin nicht gebraucht werden. Zur Herstellung der Kompilierbarkeit ohne die Einbindung unnötigen Ballasts (weiterer Drittbibliotheken) mussten sie daher in der für IRIS angepassten Zusammenstellung entfernt werden.

15.4.5.2 Modifikationen in JSPWiki (jspwiki_iisy_patches)

Im Falle der Anpassungen von JSPWiki erscheint aufgrund der sehr großen Anzahl eine vollständige Auflistung aller enthaltenen Pakete und Klassen nicht sinnvoll. Stattdessen werden hier nur die tatsächlichen Veränderungen genannt. Da die als Patches bezeichneten Änderungen in der Struktur des Teilprojektes jspwiki_iisy_patches ohnehin in eigenen Ordnern vom originalen Quellcode getrennt sind, sollte das Nachvollziehen der Änderungen auch kein allzu großes Problem darstellen.

Tabelle 20: Modifikationen am Quellcode von JSPWiki

Paket	Klasse
com.ecyrd.jspwiki	
	PageManager
	WikiContext
	WikiEngine
	WikiSession
com.ecyrd.jspwiki.auth	
	AuthenticationManager
	AuthorizationManager
	SecurityVerifier
	UserManager
com.ecyrd.jspwiki.auth.authorize	
	Group
	Role
com.ecyrd.jspwiki.auth.login	
	AbstractLoginModule
	HiddenLoginModule
	PersonDatabaseCallback
	TicketDatabaseCallback
	UserDatabaseLoginModule
	WikiCallbackHandler
com.ecyrd.jspwiki.auth.user	
	AbstractUserDatabase
	IPersonDatabase
	ITicketDatabase
	JDBCUserDatabase
	PersonDatabase

	TicketDatabase
com.ecyrd.jspwiki.diff	
	TraditionalDiffProvider
com.ecyrd.jspwiki.parser	
	JSPWikiMarkupParser
	MarkupParser
com.ecyrd.jspwiki.plugin	
	ACachedPlugin*
	Applet*
	CacheHolder*
	CachePluginManager*
	DownloadHolder*
	DownloadJobServlet*
	ImageMap*
	ImageMapServlet*
	Math*
	MathHotEnqGifWriter*
	ParameterNotFoundException*
	PluginParameterParser*
	PluginScanner*
	SbPageViewCountPlugin*
	TableOfContents
	Text2gif*
com.ecyrd.jspwiki.providers	
	CachingProvider
	FileSystemProvider
	RCSFileProvider
	VersioningFileProvider
	WikiPageProvider
com.ecyrd.jspwiki.search	
	LuceneSearchProvider
com.ecyrd.jspwiki.tags	
	BreadcrumbsTag* *
	ContentTag
	LinkToTag
	MessagesTag
	PageNameTag
	UserCheckTag
com.ecyrd.jspwiki.ui	
	AbstractCommand
	WikiCommand
com.ecyrd.jspwiki.util	
	HttpUtil
iisy.auth.impl	
	WikiTicketHandler
iisy.mvc	
	IisyWikiOp
iisy.util	
	IPConstant
	LoginService
	PasswordService
	PersonUtil

*) Diese Plugins wurden nicht von den Projektarbeitern entwickelt.

***) Nur nötig, um mit Java ab Version 5 übersetzen zu können.

15.4.5.3 Modifikationen in Regain (iisy-regain)

Tabelle 21: Modifikationen am Quellcode von Regain

Paket	Klasse
net.sf.regain	
	RegainConstants
	RegainException
	RegainToolkit
	XmlToolkit
net.sf.regain.crawler	
	Crawler
	CrawlerJob
	CrawlerToolkit
	DocWriterManager
	ErrorLogger
	HttpException
	IndexWriterManager
	Main
	Profiler
	RedirectException
	UrlChecker
	WriterManager
	WriterManagerFactory
net.sf.regain.crawler.access	
	CrawlerAccessController
net.sf.regain.crawler.config	
	AuxiliaryField
	CrawlerConfig
	DummyCrawlerConfig
	PrefixUrlMatcher
	PreparatorConfig
	PreparatorSettings
	RegexUrlMatcher
	StartUrl
	UrlMatcher
	UrlPattern
	WhiteListEntry
	XmlCrawlerConfig
net.sf.regain.crawler.config.ext	
	ExtStartUrl
	ExtUrlPattern
net.sf.regain.crawler.config.metadata	
	DocumentCorrelation
	DocumentCorrelationConfig
net.sf.regain.crawler.config.pseudo	
	AbstractPseudoDocumentConfig
	DbPseudoDocumentConfig
	FieldConfig
	IDbPseudoDocumentConfig

	IFieldConfig
	IPseudoDocumentConfig
	XmlPseudoDocumentConfig
net.sf.regain.crawler.config.writer	
	DatabaseWriterConfig
	WriterConfig
net.sf.regain.crawler.document	
	AbstractPreparator
	DbDocumentFactory
	DocumentFactory
	FieldContent
	FieldProperty
	HttpDownloadThread
	IFieldContent
	IteratorTokenStream
	PathElement
	Preparator
	PreparatorFactory
	PseudoDocument
	RawDocument
net.sf.regain.crawler.ext	
	WordlistCache
	WordlistWriter
net.sf.regain.crawler.preparator	
	AbstractJacobMsOfficePreparator
	EmptyPreparator
	ExternalPreparator
	HtmlPreparator
	IfilterPreparator
	JacobMsExcelPreparator
	JacobMsPowerPointPreparator
	JacobMsWordPreparator
	OpenOfficePreparator
	PdfBoxPreparator
	PlainTextPreparator
	PoiMsExcelPreparator
	PoiMsPowerPointPreparator
	PoiMsWordPreparator
	SimpleRtfPreparator
	SwingRtfPreparator
	XmlPreparator
net.sf.regain.crawler.preparator.html	
	AbstractExtractor
	HtmlContentExtractor
	HtmlPathExtractor
net.sf.regain.crawler.preparator.ifilter	
	IfilterWrapper
net.sf.regain.crawler.preparator.iisy	
	IisyEmptyPreparator
	IisyExternalPreparator
	IisyHtmlPreparator
	IisyIfilterPreparator
	IisyJacobMsExcelPreparator

	lisyJacobMsPowerPointPreparator
	lisyJacobMsWordPreparator
	lisyOpenOfficePreparator
	lisyPdfBoxPreparator
	lisyPlainTextPreparator
	lisyPoiMsExcelPreparator
	lisyPoiMsPowerPointPreparator
	lisyPoiMsWordPreparator
	lisyPseudoDocumentPreparator
	lisySimpleRtfPreparator
	lisySwingRtfPreparator
	lisyXmlPreparator
net.sf.regain.crawler.preparator.pseudo	
	PseudoDocumentPreparator
net.sf.regain.crawler.preparator.rtf	
	RtfFilterReader
net.sf.regain.crawler.pseudodocument	
	AbstractPseudoDocumentProvider
	DatabaseDocumentProvider
	IDatabaseDocumentProvider
	IPseudoDocumentProvider
	IXmlDocumentProvider
	PseudoDocumentException
	XmlDocumentProvider
net.sf.regain.search	
	IndexSearcherManager
	SearchConstants
	SearchToolkit
net.sf.regain.search.access	
	SearchAccessController
net.sf.regain.search.config	
	DefaultSearchConfigFactory
	IndexConfig
	SearchConfig
	SearchConfigFactory
	XmlSearchConfig
net.sf.regain.search.results	
	MergedHits
	MultipleSearchResults
	SearchResults
	SingleSearchResults
net.sf.regain.search.sharedlib	
	CheckTag
	ListTag
	MsgTag
	NavigationTag
	ParamTag
net.sf.regain.search.sharedlib.error	
	AbstractErrorTag
	MessageTag
	StacktraceTag
net.sf.regain.search.sharedlib.hit	
	AbstractHitTag

	FieldTag
	LinkTag
	PathTag
	ScoreTag
	SizeTag
	TypeiconTag
	UrlTag
net.sf.regain.search.sharedlib.input	
	FieldlistTag
	HiddenparamTag
	MaxresultsTag
	QueryTag
	SubmitTag
net.sf.regain.search.sharedlib.stats	
	FromTag
	QueryTag
	SearchtimeTag
	ToTag
	TotalTag
net.sf.regain.ui.desktop	
	BrowserSelectorFrame
	DesktopConstants
	DesktopToolkit
	FileService
	IndexUpdateManager
	Main
	TrayIconManager
net.sf.regain.ui.desktop.config	
	DesktopConfig
	XmlDesktopConfig
net.sf.regain.ui.desktop.config.sharedlib	
	EditlistTag
	FormTag
	IntervalTag
	TextTag
net.sf.regain.ui.desktop.status.sharedlib	
	CurrentindexTag
	IndexupdateTag
	IndexupdatecontrolTag
	LogTag
	ProfilerTag
net.sf.regain.ui.desktop.status.sharedlib. autoupdate	
	FormTag
	MetaTag
net.sf.regain.ui.server	
	FileServlet
net.sf.regain.ui.server.taglib	
	CheckTag
	ListTag
	MsgTag
	NavigationTag
	ParamTag

net.sf.regain.ui.server.taglib.error	
	MessageTag
	StacktraceTag
net.sf.regain.ui.server.taglib.hit	
	FieldTag
	LinkTag
	PathTag
	ScoreTag
	SizeTag
	TypeiconTag
	UriTag
net.sf.regain.ui.server.taglib.input	
	FieldlistTag
	HiddenparamTag
	MaxresultsTag
	QueryTag
	SubmitTag
net.sf.regain.ui.server.taglib.stats	
	FromTag
	QueryTag
	SearchtimeTag
	ToTag
	TotalTag
net.sf.regain.util.io	
	HtmlEntities
	Localizer
	MemoryAppender
	MultiLocalizer
	Printer
net.sf.regain.util.sharedtag	
	PageRequest
	PageResponse
	SharedTag
net.sf.regain.util.sharedtag.simple	
	Executer
	ExecuterParser
	RedirectException
	SharedTagExecuter
	SharedTagResource
	SharedTagService
	SimplePageRequest
	SimplePageResponse
	TextExecuter
net.sf.regain.util.sharedtag.taglib	
	ExtendedJspException
	JspPageRequest
	JspPageResponse
	SharedTagWrapperTag
net.sf.regain.util.ui	
	BrowserLauncher

15.5 Build Prozess

Jedes Teilprojekt enthält eine Build-Datei (build.xml) für das Werkzeug Ant (näheres hierzu siehe [Ant]). Damit erfolgt der Prozess des Übersetzens sowie der Zusammenstellung und Erzeugung von fertigen Modulen kontrolliert und automatisiert.

Wie in Anhang 15.4.2 bereits erwähnt, enthält das Projekt iisy-tools eine Build-Datei zur Erzeugung der gesamten Applikation mit allen Teilanwendungen (Webapplikation, Crawler, Quartz-Server). Dazu werden die entsprechenden Targets der jeweiligen Build-Dateien aller Teilprojekte kontrolliert nacheinander in der richtigen Reihenfolge aufgerufen. Das ist wichtig, da einige Teilprojekte aufeinander aufbauen und somit bei Änderungen am zugrunde liegenden Projekt dieses zuerst übersetzt werden muss, damit die Modifikationen auch im abhängigen Projekt enthalten sind. Abbildung 81 zeigt die Abhängigkeiten der einzelnen Teilprojekte. Alle Projekte benötigen die Basisfunktionalität aus iisy-base und laufen letztlich im Kern der Webanwendung, also in iisy-core, zusammen.

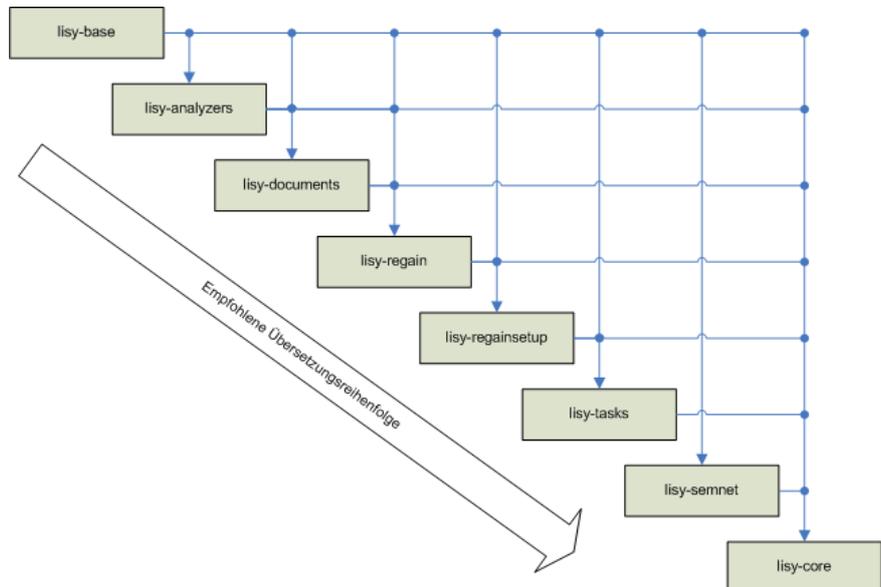


Abbildung 81: Projektabhängigkeiten

Das Target **create-all-iisy** aus der Build-Datei von iisy-tools führt einen vollständigen **create**-Prozess – alle Aspekte vom Aufräumen (**clean**), Übersetzen (**compile**), Zusammenstellen (**jar**, **dist**) bis zum Umkopieren an die eigentliche Zieladresse (**deploy**) – für alle Teilprojekte in der empfohlenen Reihenfolge durch. Die Aufgaben können für jedes Teilprojekt aber auch einzeln angestoßen werden.

16 Impressum



Univ. Prof. DI Hermann Kaufmann
Technische Universität München
Lehrstuhl Baukonstruktion, Fachgebiet Holzbau

Hochschule **Rosenheim**
University of Applied Sciences



Prof. Dr. Hartmut Ernst
Hochschule für angewandte Wissenschaften Fachhochschule Rosenheim
Fachbereich Informatik

Projektleitung TP 21

Prof. Dr. Hartmut Ernst

Projektleitung TP 22

Univ. Prof. DI Hermann Kaufmann

Projektmitarbeiter TUM

M. Sc., Dipl.-Inf. (FH), Dipl.-Ing. (FH) Johann Bellmann

Projektmitarbeiter HR

M. Eng., Dipl.-Ing. (FH) Michael Breckl-Stock
M. Sc., Dipl.-Inf. (FH) Markus Sebastian Franke
M. Eng., Dipl.-Ing. (FH) Oliver Oechsle
Dipl.-Inf. (FH) Sebastian Schneider
Dipl.-Inf. (FH) Andrea Till

Rosenheim, im Mai 2008