

Warum ist eine Neuerstellung notwendig?.....	3
Probleme mit dem ‚alten‘ Meyers-Auftritt	3
Datenmodell	3
Architektur	3
Seitengenerierung	3
URLs / Suchmaschinen.....	3
Eine neue Architektur für die „retro-Bibliothek“	4
Anforderungen.....	4
Besucher des Auftritts.....	5
Anhang: Datenmodell.....	10
Tabellen für die Bücher.....	11
Abbildungsverzeichnis	15
Verzeichnis der Tabellen	16
Stichwortverzeichnis	17

Warum ist eine Neuerstellung notwendig?

Das Meyers-Projekt¹ hatte die Zielsetzung, das Meyers Konversationslexikon von 1888 zu digitalisieren und im Internet verfügbar zu machen. Auf dieses Ziel ist das Datenmodell und der Webaufttritt ausgelegt. Im Laufe der Zeit haben sich andere Werke zum Meyers hinzugesellt. Um diese anderen Werke verfügbar zu machen, wurde das Datenmodell und der Auftritt „missbraucht“ – also nicht in der Art und Weise benutzt, für die er gedacht war.

Probleme mit dem ‚alten‘ Meyers-Aufttritt

Datenmodell

Bei dem Meyers ist das Datenmodell strikt linear aufgebaut:

Werk ← Band ← Seite ← Schlagwort

Es wäre dadurch nur über umfangreiche Joins möglich, Suchen auf ein Werk zu beschränken. Zudem enthält ein Band eine Nummer (eben die Nummer des Bandes), die *unique* (also einmalig) ist – daher haben alle anderen Werke Bandzählungen, die nicht der Wirklichkeit entsprechen.

Einige der Werke haben in sich Unterteilungen mit eigenständiger Seitennummerierung. Daher ist eine weitere Unterteilung eines Bandes in „Teile“, die eigenständige Seitennummerierungen haben, wünschenswert.

Architektur

Das Meyers basiert auf Servlets in einem Tomcat-Container, die über einen selbsterstellten Abstraktionslayer² auf die Datenbank zugreifen und direkt HTML-Code für die Antwort an den User erstellen. Die Vermischung von Java-Code und HTML macht die Wartung der Anwendung schwierig.

Seitengenerierung

Die Servlets generieren HTML 4.1-Transistional-Code, das Layout ist über verschachtelte Tabellen erstellt. Eine Trennung von Inhalt und Darstellung ist nur schwer möglich. Die textuellen Inhalte der retrodigitalisierten Seiten haben zwar Stichworte und Verweise, andere Spezialitäten im Text, wie *S p e r r d r u c k*, *Kursivdruck* oder Tabellen lassen sich aber nicht realisieren.

URLs / Suchmaschinen

Der Aufbau der verwendeten URLs ist bereits für Suchmaschinen optimiert. Allerdings läuft die Anwendung unter dem Port 8080 – was einerseits einige User ausschließt, die diesen Port in der Firewall gesperrt haben, andererseits bei Analysetools wie z. B. Google Analytics eine Bildung von Channels ausschließt. Der Hostname ist zur Zeit „susi.e-technik.uni-ulm.de“, was bei Suchmaschinen eine große Gewichtung auf „Universitäten“ bewirkt.

¹ <http://www.meyers-konversationslexikon.de>

² <http://www.dlmeta.de>

Eine neue Architektur für die „retro-Bibliothek“

Anforderungen

Die Anforderungen an eine neue Architektur gliedern sich nach den verschiedenen Bedürfnissen der Anwendergruppen:

- **Besucher** des Auftritts (Suchen & Finden, Stöbern, Hinweise auf Fehler geben etc.)
- **Korrektoren** (Handhabung der Seiten, Korrekturhilfen etc.)
- **Administratoren** (Fehlersuche, Wartung, Backup, Indexerstellung etc.)
- **Fremde (Web-)Administratoren** (Verlinkung, Schnittstellen etc.)
- **Programmierer** (Wartung des Codes, modularer Aufbau etc.)
- **Suchmaschinen** (Indizierbarkeit, Sitemaps etc.)

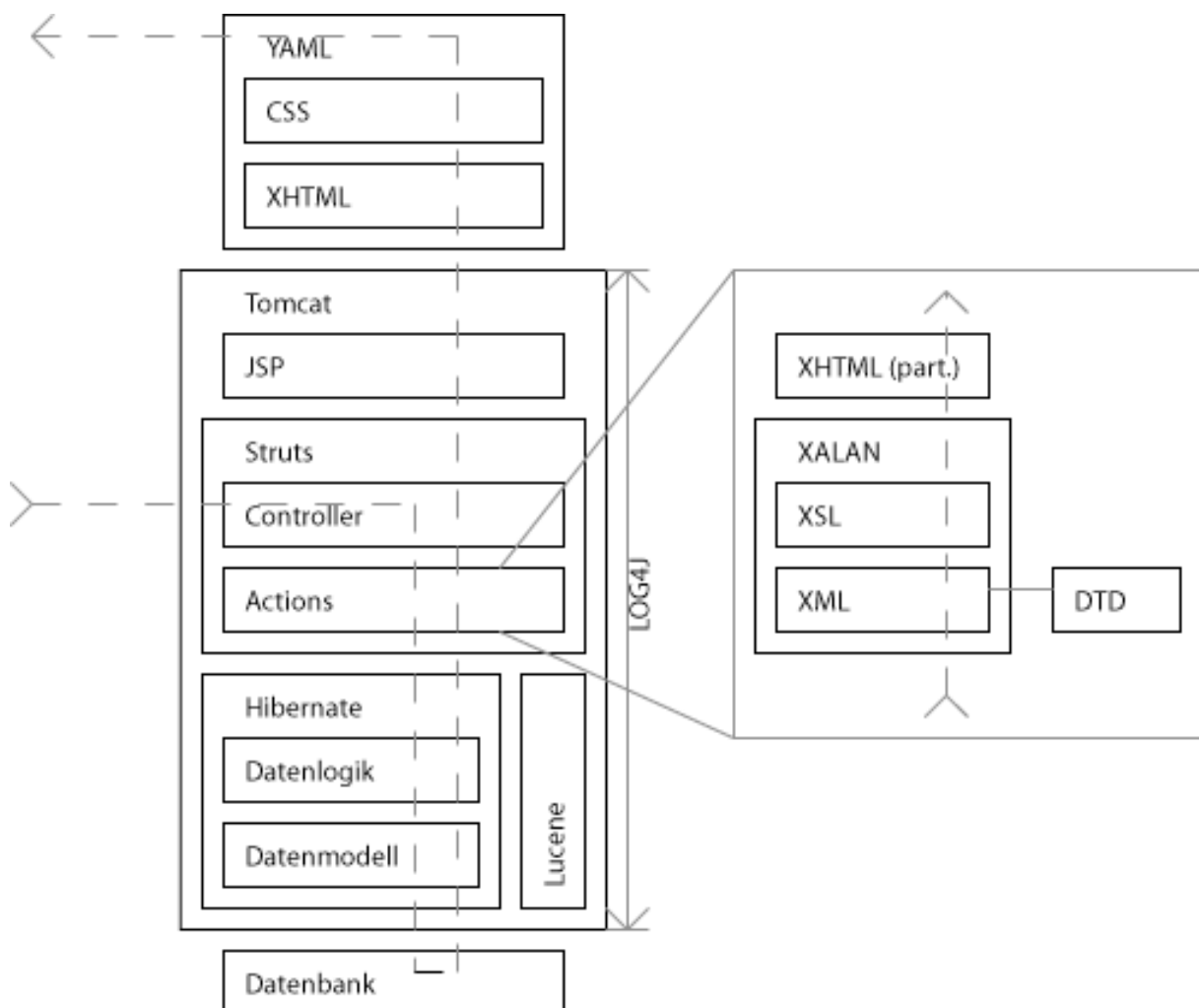


Abbildung 1 – Architekturübersicht.

YAML	Yet Another Multicolumn Layout
CSS	Cascading Style Sheets
XHTML	HTML auf der Basis von XML
XALAN	XSLT (Extensible Stylesheet Language Transformation) processor
XSL	Extensible Stylesheet Language
XML	Extensible Markup Language

DTD	Document Type Definition
Tomcat	Application Server
JSP	Java Server Pages
Struts	Model 2 Controller
Lucene	Volltextsuchmaschine
Hibernate	Datenbankpersistenz und -abstraktions Layer
Log4j	Logging-Modul für Java

Tabelle 1 - Erläuterung der Abkürzungen.

Besucher des Auftritts

85% der Besucher kommen von Google direkt auf eine Seite eines Werkes. Die verwendete Bildschirmauflösung liegt größtenteils bei 1025*768 oder höher (siehe Tabellen unten). Die Seite eines Werkes sollte bei einem solchen Zugriff alle wesentlichen Informationen (Werk, Band etc., aber auch die „Art“ der Information) auf einen Blick zeigen.

Der „retro|bib“-Schriftzug in Fraktur weist dabei auf alte Quellen hin. Werkname, Autor, Verlag und Erscheinungszeitpunkt zusammen mit dem Band sind sofort ersichtlich.

Die Teile des Auftritts und die Meta-Daten sind in serifenloser Schrift und zur Abgrenzung und besseren Lesbarkeit der Text der Seite in einer Schriftart mit Serifen gesetzt.

The screenshot shows a web browser window displaying the 'retro|bib' website. The page title is 'retro|bib Die Retro-Bibliothek'. The main content area features a search bar and a list of navigation links. The current page is the entry for 'Wässer, abgezogene' from the 'Meyers Konversationslexikon', 16th volume, Uralsk - Zz. The entry text describes the process of softening hard water using gypsum and carbon dioxide. The page also includes a sidebar with 'Google-Anzeigen' and 'Die eigene Sauna'.

Abbildung 2 - Seite.html bei 1024*768 Pixeln Auflösung.

Da die Seite in XHTML gesetzt ist und zur Darstellung CSS verwendet, lässt sie sich auch mit Screenreadern ohne Probleme darstellen. Für Screenreader werden dabei ganz am Anfang zwei Links sichtbar: „Skip to the navigation“ und „Skip to the content“. Benutzer eines Screenreaders können damit schneller zu den wichtigen Teilen der Seite navigieren.



Abbildung 3 - Darstellung der Seite mit Lynx.

Quellen	Besuche	% Zugriffe
google (organic)	169552	0.85
de.wikipedia.org (referral)	6431	0.03
(direct) ((none))	5543	0.03
suche.t-online.de (referral)	3823	0.02
meyers-konversationslexikon.de (referral)	2606	0.01

Tabelle 2 - Quellen der Besucher (Stand 19.10.2007; Zeitraum: 1 Monat).

Bildschirmauflösung	Besuche
1024*768	96144
1280*1024	45557
1280*800	20841
800*600	8108
1152*864	7380
1440*900	5523
1680*1050	3720
1280*960	2553
1400*1050	2338
1600*1200	1884

Tabelle 3 - Bildschirmauflösungen (Stand 19.20.2007; Zeitraum: 1 Monat).

Ein zweiter wichtiger Zugang zu den Daten in der retro|bib ist die Suche. Auf jeder Seite ist über dem Hauptmenü links ein Suchfeld, mit dem eine „Schnellsuche“ durchgeführt werden kann. Schnellsuche meint hier eine Suche ohne eine Eingrenzung auf bestimmte Werke. Im Gegensatz dazu ist auf der Seite „suche.html“ die Eingrenzung auf Werke möglich.



Abbildung 4 - suche.html.

Die Suche ist über einen Volltextindex realisiert. Zusätzlich zu dem Inhalt der Seiten werden Treffer in den Schlagworten höher gewichtet. Werden keine Treffer gefunden, oder liegt die Relevanz des ersten Treffers unter 10%, wird ein alternatives Suchwort vorgeschlagen. In einem Vorschaubereich wird das beste Textfragment der Seite mit hervorgehobenen Suchwörtern angezeigt.

The screenshot shows a web browser window with the URL <http://www.retrobibliothek.de/retro|bib/suchtreffer.html?suchtext=Wasseralfingen>. The page title is "retro|bib Die Retro-Bibliothek". The search results are for "Wasseralfingen" and show 13 results. The first result is "100% Meyers → 16. Band: Uralsk - Zz → Hauptstück: 0416 Wässer, abgezogene - Wasserbau" with an "Öffnen" link. The second result is "1% Meyers → 16. Band: Uralsk - Zz → Hauptstück: 0775 Württemberg (Bergbau, Industrie)" with an "Öffnen" link. The third result is "1% Meyers → Schlüssel → Schlüssel: 0065 Geographie: Liechtenstein. Oesterreich-Ungarn" with an "Öffnen" link. The fourth result is "1% Kuenstler → Hauptstück → Lexikon: 0313 König - Kornek" with an "Öffnen" link. The fifth result is "1% Meyers → 19. Band: Jahres-Supplement 1891[...] → Hauptstück: 0109 Bergbahnen (Zahnradbahnen)" with an "Öffnen" link. On the left side, there is a "Schnellsuche:" field and a "Los!" button. Below that is a "Hauptmenü" with links for "Hauptseite", "Suchen", "Stöbern", "Technisches", "zum Meyers", "Autorennamen", "Abkürzungen", "Rundgang", "für Korrektoren", "Fragen & Antworten", and "Korrekturhilfe". On the right side, there are "Google-Anzeigen" for "Herz Geschenkkidee", "Die eigene Sauna", "Asics Schuhe & Bekleidung", and "Marken Sicherheitsschuh".

Abbildung 5 - suchtreffer.html mit alternativen Suchwörtern.

Der dritte Zugang zu den Daten erfolgt über das „Stöbern“ durch den Datenbestand. Hierbei klickt sich der Benutzer von oben (dem Werk) nach unten (die Seite) durch.

retro|bib - In der Bibliothek stöbern

http://www.retrobibliothek.de/retrobib/stoebn.html

1. Raiba 2. MKL 3. MKL (local) 4. r|b 5. r|b (lokal) 6. Wiki Marc Developer Eigene News Wiki Amazon Google Uni W3C Validation

retro|bib - In der Bibliothe...

Impressum | Login für Korrektoren

retro | bib

Die Retro-Bibliothek

Schnellsuche:

Los! Löschen

Hauptmenü

- Hauptseite
- Suchen
- Stöbern

Technisches

- Technik
- Statistik

zum Meyers

- Autorennamen
- Abkürzungen
- Rundgang

für Korrektoren

- Fragen & Antworten
- Korrekturhilfe

In den Werken Stöbern

Hier können Sie durch die vorhandenen Werke wie in einer Bibliothek stöbern bzw. blättern. Klicken Sie zur Auswahl auf die jeweiligen Titel.

 **Meyers Konversationslexikon**

Autorenkollektiv, VERLAG DES BIBLIOGRAPHISCHEN INSTITUTS, LEIPZIG UND WIEN, Vierte Auflage, 1885-1892

 **Brockhaus' Konversationslexikon**

Autorenkollektiv, F. A. BROCKHAUS IN LEIPZIG, BERLIN UND WIEN, 14. Auflage, 1894-1896

 **Merck's Warenlexikon**

Autorenkollektiv, VERLAG VON G. A. GLOECKNER, LEIPZIG, Dritte Auflage, 1884

 Beschreibung der im Handel vorkommenden Natur- und Kunsterzeugnisse unter besonderer Berücksichtigung der chemisch-technischen und anderer Fabrikate, der Drogen- und Farbewaren, der Kolonialwaren, der Landesprodukte, der Material- und Mineralwaren.

 **Handbuch der Drogisten-Praxis**

Google-Anzeigen

[Herz Geschenkkidee](#)
Hier finden Sie schöne interessante Geschenkkideen & sparen bis zu 75%!
[www.Preisvergleich.de/](#)

[Die eigene Sauna](#)
Spitzen-Qualität zu Werkspreisen Jetzt kostenlosen Katalog anfordern
[www.Hofmann-Sauna.c](#)

[Asics Schuhe & Bekleidung](#)
in riesiger Auswahl, mit bis zu 33% Rabatt - hier kaufen Sie günstig!
[www.Sportpoint24.com](#)

[Marken Sicherheitsschuh](#)

Abbildung 6 - stoebn.html

Anhang: Datenmodell

Das Datenmodell dient der Verwaltung und inhaltlichen Speicherung der Seiten der Werke in einer relationalen Datenbank. Jede Tabelle, und daher nicht jedes Mal erläutert, enthält eine Id, ein Timestamp (der Zeitpunkt der letzten Änderung) und ein Löschflag (Deleted). Timestamp und Löschflag dienen der Synchronisation, sprich dem Backup. Alle Daten werden als UTF-8 gespeichert.

Tabelle Account			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
eMail	Varchar	Unique	
Password	Varchar		MD5-Hash des Passwortes

Tabelle 4 - Tabelle Account

In der Tabelle Account werden die Zugangsdaten der Benutzer gespeichert. Ein Benutzer identifiziert sich am System über seine eMail (als Login) und ein Passwort, welches als MD5-Hash gespeichert wird.

Tabelle Accountsetting			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid		Fremdschlüssel	
Name	Varchar		Name der Einstellung
Value	Varchar		Wert der Einstellung

Tabelle 5 - Tabelle Accountsetting

Da jeder Benutzer individuelle Einstellungen haben kann (und diese im Verlaufe der Entwicklung des Projektes zunehmen können) werden sie in einer eigenen Tabelle gespeichert.

Tabelle Globalsetting			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Name	Varchar		Name der Einstellung
Value	Varchar		Wert der Einstellung

Tabelle 6 - Tabelle Globalsetting

Generelle Einstellungen der Anwendung werden in der Tabelle Globalsetting gespeichert.

Tabellen für die Bücher

In einer groben Aufteilung ist die Struktur Werk ← Band ← Teil ← Seite angedacht. Wie oben erwähnt, dient der Teil der Unterteilung eines Bandes in Abschnitte mit eigener Seitenzählung.

Tabelle Werk			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Sort	Varchar		Sortierreihenfolge in Listen
Name	Varchar		Name des Werkes
Shortname	Varchar		Kurzname des Werkes, z. B. für URLs
Publisher	Varchar		Verlag des Werkes
Author	Varchar		Autor des Werkes
Issued	Varchar		Erscheinungsjahr/-ort des Werkes
Description	Text		Weitere Informationen zu dem Werk

Tabelle 7 - Tabelle Werk

Die Tabelle Werk beschreibt ein Werk in dem Sinne einer zusammengehörigen und abgeschlossenen Sammlung von Bänden, wie z. B. beim Meyers die 19 Bände. Auch einzelne Bücher, wie z. B. das Mercks, werden immer auch als Werk erfasst.

Tabelle Band			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Werkid	Integer	Fremdschlüssel	Das Werk, zu dem der Band gehört
Sort	Varchar		Sortierreihenfolge in Listen (hier in dem Werk)
Name	Varchar		Name des Bandes
Publisher	Varchar		Verlag des Bandes
Author	Varchar		Autor des Bandes
Issued	Varchar		Erscheinungsjahr/-ort des Bandes
Description	Text		Weitere Informationen zu dem Band

Tabelle 8 - Tabelle Band

Die Tabelle Band scheint auf den ersten Blick redundante Informationen bezüglich des Werkes zu haben. Das Feld `Issued` z. B. kann allerdings innerhalb von Werken verschieden

sein. Bei angedachten Sammlungen, wie z. B. Dissertationen um 1900 (von denen ich etwa 100 zu digitalisieren gedenke) sind auch die Autoren von Band zu Band unterschiedlich.

Tabelle Teil			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Bandid	Integer	Fremdschlüssel	Der Band, zu dem der Teil gehört
Sort	Varchar		Sortierreihenfolge in Listen (hier in dem Band)
Visible	Boolean		Soll der Teil angezeigt werden, z. B. als Kapitel? Oder ist er nur eine unsichtbare Seitenzahlunterteilung?
Name	Varchar		Name des Bandes
Description	Text		Weitere Informationen zu dem Band

Tabelle 9 - Tabelle Teil

Teile zerteilen Bände in einzelne Abschnitte, wie z. B. in die Einleitung, die römisch gezählt wird, den Hauptteil, der neu arabisch zählt und den Anhang, der wiederum eine neue Seitenzählung haben kann.

Tabelle Seite			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Teilid	Integer	Fremdschlüssel	Der Teil, zu dem die Seite gehört
Sort	Varchar		Sortierreihenfolge in Listen (hier in dem Teil)
Pagenumber	Varchar		Seitennummer (als Varchar, damit z. B. römische Zahlen gehen)
Startword	Varchar		Von Stichwort...
Endword	Varchar		... bis Stichwort
Status	Integer		Der Status der Seite (unkorrigiert, normal, fein, tagged)
Blockmarker	Boolean		Ist die Seite ausgeklinkt?
Hash	Varchar		MD5-Hash des Inhalts, für Konsistenzprüfungen

Tabelle 10 - Tabelle Seite

Die Seite ist das Herzstück der Datenbank und enthält die Metadaten zu den Textdaten.

Tabelle Seitentext			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Seiteid	Integer	Fremdschlüssel	Die Seite, die dieser Faksimile darstellt
Content	Clob		Der Text
Encoding	Varchar		Das Encoding des Textes (UTF-8 etc.)
Hash	Varchar		MD5-Hash des Inhalts, für Konsistenzprüfungen

Tabelle 11 - Tabelle Seitentext

Die Tabelle Seitentext enthält die eigentlichen Texte der Seiten. Durch diese Trennung ist es denkbar, viele Text (Versionisierung) zu einer Seite zu haben.

Tabelle Faksimile			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Seiteid	Integer	Fremdschlüssel	Die Seite, die dieser Faksimile darstellt
Scan	Blob		Das Bild
Mimetype	Varchar		Die Art des Bildes (gif/jpeg etc.)
Hash	Varchar		MD5-Hash des Inhalts, für Konsistenzprüfungen

Tabelle 12 - Tabelle Faksimile

Die Tabelle Faksimile enthält die Bildansicht (den Scan) zu einer Seite.

Tabelle Hiresfaksimile			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Seiteid	Integer	Fremdschlüssel	Die Seite, die dieser Faksimile darstellt

Scan	Blob	Das Bild
Mimetype	Varchar	Die Art des Bildes (gif/jpeg etc.)
Hash	Varchar	MD5-Hash des Inhalts, für Konsistenzprüfungen

Tabelle 13 – Hiresfaksimile

Die Tabelle Hiresfaksimile speichert hochaufgelöste Scans z. B. für Chromlithographien.

Tabelle Schlagwort			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Seiteid	Integer	Fremdschlüssel	Die Seite, die dieser Faksimile darstellt
Keyword	Varchar		Das Schlagwort

Tabelle 14 - Tabelle Schlagwort

In der Tabelle Schlagwort werden (im Prinzip redundant) die Schlagworte einer Seite gehalten. Dieser „Kunstgriff“ ermöglicht die sinnvolle Quervernetzung innerhalb der Werke mit den Verweisen im Text (s. d.; siehe Stichwort etc.).

Tabelle Siehewort			
Spaltenname	Datentyp	Art	Beschreibung
Id	Integer	Primärschlüssel	
Oldid	Integer		Für die Migration der Daten
Timestamp	Timestamp		Letzte Änderung für die Synchronisation
Deleted	Boolean		
Accountid	Integer	Fremdschlüssel	Der Account, der die letzte Änderung durchgeführt hat
Seiteid	Integer	Fremdschlüssel	Die Seite, die dieser Faksimile darstellt
Seeword	Varchar		Das Verweiswort

Tabelle 15 - Tabelle Siehewort

Mit Hilfe der Tabelle Siehewort wird wir bei den Schlagwörtern eine Quervernetzung möglich.

Abbildungsverzeichnis

Abbildung 1 – Architekturübersicht.....	4
Abbildung 2 - Seite.html bei 1024*768 Pixeln Auflösung.....	5
Abbildung 3 - Darstellung der Seite mit Lynx.....	6
Abbildung 4 - suche.html.....	7
Abbildung 5 - suchtreffer.html mit alternativen Suchwörtern.....	8
Abbildung 6 - stoebern.html	9

Verzeichnis der Tabellen

Tabelle 1 - Erläuterung der Abkürzungen.....	5
Tabelle 2 - Quellen der Besucher (Stand 19.10.2007; Zeitraum: 1 Monat).....	6
Tabelle 3 - Bildschirmauflösungen (Stand 19.20.2007; Zeitraum: 1 Monat).....	6
Tabelle 4 - Tabelle Account.....	10
Tabelle 5 - Tabelle Accountsetting	10
Tabelle 6 - Tabelle Globalsetting	10
Tabelle 7 - Tabelle Werk	11
Tabelle 8 - Tabelle Band.....	11
Tabelle 9 - Tabelle Teil.....	12
Tabelle 10 - Tabelle Seite	12
Tabelle 11 - Tabelle Seitentext	13
Tabelle 12 - Tabelle Faksimile.....	13
Tabelle 13 – Hiresfaksimile	14
Tabelle 14 - Tabelle Schlagwort	14
Tabelle 15 - Tabelle Siehewort	14

Stichwortverzeichnis

Fehler! Keine Indexeinträge gefunden.