Inhaltsverzeichnis

Kurzfassung iii

Abstract v

Inhaltsverzeichnis vii

1 Einleitung 1
1.1 Problemstellung und Einordnung der Arbeit . . . 1
1.2 Lösungsansatz . . . 7
L3 Verwandte Arbeiten . . . S
1.3.1 Datenbank-Suchmaschinen . . . 9
1.3.2 Schlüsselwortbasierte Suche in Datenbanken . . . 9
1.3.2.1 Graphenbasierte Ansätze . . . 10
1.3.2.2 Ansätze basierend auf der Generierung von SQL-Anfragen 12
1.3.3 Auffinden von Datenquellen im Internet . . . 16
1.4 Zielstellung . . . 17
1.5 Gliederung der Arbeit . . . 19

2 Theoretische Grundlagen 21
2.1 Information Retrieval . . . 21
2.1.1 Boolesches Retrieval . . . 23
2.1.1.1 Implementierung des Booleschen Retrievals . . . 25
2.1.2 Vektorraum-Modell . . . 25
2.1.2.1 Gewichtungsmethoden für Textdokumente . . . 28
2.1.2.2 Implementierung des Vektorraum-Retrievals . . . 30
2.1.3 Auswahl von Datenquellen auf der Basis des Vektorraiim-Modells . 31
2.2 Relationale Datenbanken . . . 35

2.2.1 Grundlagen . . . 35
2.2.2 Relationenalgebra . . . 36
2.2.3 Punktionale und mehrwertige Abhängigkeiten . . . 37
2.2.4 Normalformen . . . 38
2.2.5 Abhängigkeits- und Verbundtreue . . . 38
2.2.6 Projection-Join-Normalform . . . 39
2.3 Zyklische und azyklische Schemata . . . 41
2.4 Universalrelationen-Modell . . . 46
2.4.1 Grundlegende Annahmen . . . 48
2.4.2 Fensterfunktionen mit denotationaler Semantik . . . 50
2.4.3 Fensterfunktionen mit operationaler Semantik . . . 52

3 Grundlagen und Diskussion der Universalrelationen-Schnittstelle 57
3.1 Tableaus und deren Anwendung . . . 57
3.1.1 Optimierung unter schwacher Äquivalenz . . . 58
3.1.2 Ableitung von mehrwertigen Abhängigkeiten . . . 63
3.2 Umsetzung der Fensterfunktion . . . 67
3.2.1 Anfrageinterpretation bei a-azyklischen Objektschemata . . . 68
3.2.2 Graham-Reduktion mit geschützten Knoten . . . 69
3.2.3 Probleme bei Zyklen . . . 70
3.2.4 Anfrageinterpretation mit maximalen Objekten . . . 73
3.3 Automatische Bestimmung maximaler Objekte . . . 76
3.4 Objekte vs. Relationen . . . 79

4 Datenmodell für den Zugriff auf eine Web-Datenbank &; Konzeption der Datenbank-Suchmaschine 83
4.1 Datenmodell . . . 83
4.1.1 Grundlagen . . . 84
4.1.2 IR-Anfragen und strukturierte Anfragen . . . 90
4.1.3 3-Ebenen-Modell für den Zugriff auf Web-Datenbanken . . . 92
4.1.4 Optimierung des Datenbankschemas auf Objekt-Ebene &; Ermittlung des Projektionsschemas . . . 93
4.1.5 Erweiterung des Datenmodells zur Behandlung zyklischer Datenbankschemata und Verbundabhängigkeiten . . . 95
4.2 Überblick über die Arbeitsweise der Datenbank-Suchmaschine . . . 96

4.3 Technische Grundlagen der Suchmaschine . . . 99
4.3.1 Web-Suchmaschinen . . . 99
4.3.2 Web-Services . . . 100
4.4 Architektur der Datenbank-Suchmaschine . . . 104
4.5 Erfassung der notwendigen Informationen für den Zugriff auf eine Web-Datenbank . . . 108

5 Verarbeitung von IR-Anfragen durch die Suchmaschine 117
5.1 Vektorraum-Retrieval . . . 118
5.1.1 Kriterien für strukturierte Anfragen . . . 118
5.1.2 Ableitung strukturierter Anfragen für eine Datenbank . . . 119
5.1.3 Bewertung von Datenbanken auf der Basis strukturierter Anfragen . 121
5.1.4 Algorithmus zur Ableitung der strukturierten Anfragen für eine Web-DaLenbank . . . 124
5.1.5 Berechnung der relevanten Datenbanken . . . 132
5.2 Boolesche IR-Anfagen . . . 133
5.2.1 Ableitung und Bewertung von strukturierten Anfragen im Booleschen Modell . . . 133
5.2.2 Ermittlung der relevanten Datenbanken . . . 135
5.3 Indexstrukturen . . . 136

6 Indexierung und Verarbeitung strukturierter Anfragen durch die Datenbank 141
6.1 Indexierung der Inhalte einer Web-Datenbank . . . 141
6.1.1 Indexierung atomarer Attribute . . . 142
6.1.2 Indexierung mengenwertiger Attribute . . . 145
6.2 Ableitung der SQL-Anfragen aus den strukturierten Anfragen . . . 146
6.3 Verarbeitung der SQL-Anfragen . . . 152
6.4 Nachbearbeitung im Vektorraum-Modell . . . 153
G.5 Nachbearbeitung im Booleschen Modell . . . 15G

7 Evaluierung verschiedener Teilaspekte der Datenbank-Suchmaschine 159
7.1 TPC-H-Datenbank . . . 160
7.2 Indexierimgsdauer und -umfang . . . 162
7.3 Zeitaufwand für die Verarbeitung strukturierter Anfragen durch eine Web-Datenbank . . . 164

7.4 Zeitbedarf für die Ableitung der strukturierten Anfragen . . . 167
7.5 Einfluss der verschiedenen Indexvarianten . . . 168

8 Zusammenfassung und Ausblick 171
8.1 Zusammenfassung . . . 171
8.2 Ausblick . . . 174

Literaturverzeichnis 177

Abbildungsverzeichnis 189

Tabellenverzeichnis 193

Anhang 195

A Laufende Beispiele 197
A.1 ER-Modelle . . . 197
A.1.1 ER-Modell der Bibliotheksdatenbank . . . 197
A.1.2 ER-Modell der Vorlesungsdatenbank . . . 198
A.2 Datenbankschemata . . . 199
A.2.1 Datenbankschema der Bibliotheksdafcenbank . . . 199
A.2.2 Datenbankschema der Vorlesungsdatenbank . . . 199
A.3 Konfiguration der Beispieldatenbanken . . . 199
A.3.1 Bibliotheksdatenbank . . . 199
A.3.2 Vorlesungsdatenbank . . . 200
A.4 Hypergraphen für die Datenbankschcinata auf Objekt-Ebene . . . 202
A.4.1 Bibliotheksdatenbank . . . 202
A.4.2 Voiiesungsdatenbank . . . 202
A.5 Relationale Repräsentation . . . 203
A.5.1 Bibliotheksdatenbank . . . 203
A.5.2 Entities der Bibliotheksdatenbank . . . 203
A.5.3 Vorlesungsdatenbank . . . 204
A.5.4 Entities der Vorlesungsdatenbank . . . 204
A.5.5 Invertierter Index . . . 206

B Web-Service-Dokumente 209
B.l SOAP-Nachrichten . . . 209
B.l.l SO AP-Nachricht, eingebettet in eine HTTP-P OST-Anforderung . . 209
B.1.2 SOAP-Nachricht, eingebettet in eine HTTP-Antwart . . . 209
B.2 UDDI-Dokumente . . . 210
B.2.1 businessEntity . . . 210
B.2.2 businessService . . . 211
B.2.3 tModel für Integrationsdienst . . . 211
B.2.4 bindingTemplate . . . 212
B.3 WS-Inspection-Dokument . . . 212
B.3.1 WSDL-Beschreibung des Integrationsdienstes . . . 213
B.3.2 Signatur des Dienstes . . . 213
B.3.2.1 Operationen . . . 213
B.3.2.2 Datenstrukturen . . . 213
B.3.3 WS DL-Dokument für die Dienstbeschreibung . . . 215
B.3.4 WSDL-Dokument eines Anbieters . . . 218
B.4 Anfragen an UDDI-Verzeichnisse . . . 219
B.5 Verweis auf ein WS-Inspectioii-Dokument innerhalb einer HTML-Seite . . 220

C DTD und Beispiel für die Datendefinition zum Zugriff auf eine Web-Datenbank 221
C.l DTD der Dokumente für die Datendefinition . . . 221
C.2 Beispieldokument für die Bibliotheksdatenbank . . . 222

D Unterstützung von SQL-Anfragen ohne Beachtung der Groß-/Kleinschreibung am Beispiel von Oracle und DB2 225

PPN: 255235852
Titel: Techniken für Suchmaschinen zum Auffinden relevanter Informationseinheiten in Web-Datenbanken / Gunnar Weber. - Berlin : Logos-Verl., 2006
ISBN: 3-8325-1254-3Pb.EUR 40.50, EUR 41.60 (AT), sfr 72.10
Bibliographischer Datensatz im SWB-Verbund