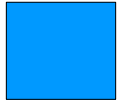


Suche im Online Katalog

Oracle Text SID Frankfurt/Main
22.02.2007

Egon Ritter

eritter@ci-database.de



Anwendungen mit Oracle Text

➤ Adressensuche

- CRM-Systeme
- Abgleich mit Sanktionslisten

➤ Dokumentensuche

- CMS-Systeme
- Fernsehsendungen

➤ Produktsuche

- Online Kataloge

Online Katalog

Katalogstart Möbelbeschläge, Möbelgriffe Baubeschläge Montagetechnik

Stichwort/ArtikelNr

[Erweiterte Produktsuche](#)

Treffer in Kategorie Katalogstart → [Erweiterte Suche](#) → [Stichwortsuche](#)


Baubeschläge (165)

- Schutzbeschläge und Türgriffe für Abschlusstüren (65)
 - Türgriffe für Abschlusstüren (48)
 - FSB (48)
 - Schutzbeschläge für Abschlusstüren (17)
 - FSB (17)
- Fenstergriffe (15)
 - FSB (15)
 - Edelstahl (14)
 - AluGrau® (1)
- Schiebetürbeschläge (12)
 - Muschelgriffe (12)
 - Edelstahl (12)
- Türdrücker-Garnituren für den Wohnbereich (55)

Trefferliste Stichwortsuche

Ihre Suche ergab 167 Treffer in 79 verschiedenen Produkten.

sortiert nach:

Abbildung	Produkt / Artikel	Marke / Modell	Details
Baubeschläge → Schutzbeschläge und Türgriffe für Abschlusstüren → Türgriffe für Abschlusstüren → FSB → Edelstahl			
	Edelstahl gefunden in Artikeln: → FSB Ros. Halb.Grt.FH 9mm_HAU [903.32.232]		→ Produktseite

Online Katalog

The screenshot shows a web interface for selecting door handle components. At the top, there are navigation tabs: 'Katalogstart', 'Möbelbeschläge, Möbelgriffe', 'Baubeschläge' (selected), and 'Montagetechnik'. Below the tabs is a search bar with the text 'Stichwort/ArtikelNr' and a search icon. A breadcrumb trail reads: 'Katalogstart > Erweiterte Suche > Auswahlassistenten > Baubeschläge > Türdrücker-Garnituren für den Wohnbereich > FSB > Edelstahl'. The main content area is titled 'Auswahlassistent FSB, Edelstahl' and shows '150 gefundene Artikel'. A button 'Artikel anzeigen' is present. Below this is a table with columns 'Artikelmerkmal', 'Auswahlmöglichkeiten', and 'Auswahl'. The table has three rows: 'Modellreihe' with a dropdown set to 'Alle' and an 'OK' button; 'Oberfläche' with the value 'matt' and a red 'X' button; and 'Türdicke' with a dropdown set to 'Alle' and an 'OK' button. A dropdown menu for 'Türdicke' is open, showing options: 'Alle', '36 - 45 (4)', '37 - 41 (58)', '38 - 41 (18)', and '38 - 42 (4)'. A note below the table says: 'Hinweis: Um ein Suchkriterium zu ändern, klicken Sie bitte auf den Button "OK" und wählen Sie die gewünschte Auswahlmöglichkeit aus der Liste und klicken Sie auf den Button "OK".' On the right side, there is a 'Ihr Warenkorb' section with the message 'Kein Artikel im Warenkorb' and links 'einschauenbestellen' and 'Direkt eingabe'. Below that is a 'Login' section with a 'Mithbenutzer' button, a text box for 'Kundennummer', another text box for '(PIN)', and a 'Bitte loggen Sie sich ein' button. Links for 'PIN anfordern', 'PIN vergessen', and 'EDV-Support' are also present. At the bottom right, there is an 'Auftragsstatus' section with the text 'Login erforderlich'.

Katalogstart Möbelbeschläge, Möbelgriffe **Baubeschläge** Montagetechnik

Stichwort/ArtikelNr

Erweiterte Produktsuche

Katalogstart > Erweiterte Suche > Auswahlassistenten > Baubeschläge > Türdrücker-Garnituren für den Wohnbereich > FSB > Edelstahl

Ihr Warenkorb

Kein Artikel im Warenkorb

[einschauenbestellen](#)

[Direkt eingabe](#)

Login **Mithbenutzer**

Einloggen und profitieren.

(PIN)

Bitte loggen Sie sich ein

[PIN anfordern](#)

[PIN vergessen](#)

[EDV-Support](#)

Auftragsstatus

Login erforderlich

Auswahlassistent FSB, Edelstahl

[←← Auswahlassistent verlassen](#)

150 gefundene Artikel.

Artikelmerkmal	Auswahlmöglichkeiten	Auswahl
Modellreihe	Alle	<input type="button" value="OK"/>
Oberfläche	matt	<input type="button" value="X"/>
Türdicke	Alle	<input type="button" value="OK"/>

Hinweis: Um ein Suchkriterium zu ändern, klicken Sie bitte auf den Button "OK" und wählen Sie die gewünschte Auswahlmöglichkeit aus der Liste und klicken Sie auf den Button "OK".

Online Katalog

- **Edelstahl**
 - Einzelteile
 - Aluminium
 - AluGrau®
 - Montagezubehör
 - Hevi
 - Hoppe
 - Jado
 - Schetter
 - VellüVell
 - Sonstige
-
- Türdrücker-Garnituren für den Objektbereich und für Rohrrahmentüren
 - Schutzbeschläge und Türgriffe für Abschlussluren
 - Schließtechnik
 - Elektronisches Schließsystem Dialock
 - Türbänder
 - Türschließer
 - Dichtungen und Schwellen
 - Schiebetürbeschläge
 - Glastürbeschläge
 - Sanitärbeschläge
 - Fenstergriffe

58 gefundene Artikel.

→ Artikel anzeigen

Artikelmerkmal	Auswahlmöglichkeiten	Auswahl
Modellreihe	Alle	OK
Oberfläche	matt	X
Türdicke	37 - 41	X

Hinweise: Um ein Suchkriterium einzustellen wählen Sie den gewünschten Wert in der Auswahlliste und klicken Sie bitte auf den Button "OK". Mit den Button "X" können Sie die Auswahl wieder aufheben.

↳ Artikel anzeigen

Trefferliste

Ihre Suche ergab 58 Treffer in 18 verschiedenen Produkten.

Abbildung	Produkt / Artikel	Marke / Modell	Details
	Baubeschläge → Türdrücker-Garnituren für den Wohnbereich → FSB → Edelstahl		
	Modell 1126/1731/1735 gefunden in Artikeln:		→ Produkseite
	→ EB-Garnitur Ø Edelstahl matt 1902.91.0500		
	→ ZVL-Garnitur Ø Edelstahl matt r 1902.95.0500		
	→ PZ-Garnitur Ø Edelstahl matt 1902.92.0500		
	→ ZVL-Garnitur Ø Edelstahl matt l 1902.95.0500		

Login **Mitbenutzer**

Einloggen und profitieren

Kundennummer

(PIN)

Bitte loggen Sie sich ein [→](#)

→ [PIN anfordern](#)

→ [PIN vergessen](#)

→ [EDV-Support](#)

[Antragsstatus](#)

Login erforderlich

[Ihre Merkzettel](#)

Login erforderlich

[CD-Bestellung](#)

Login erforderlich

[Seite drucken](#)

[Weiterempfehlen](#)

[zu Favoriten](#)



Agenda

- Anforderungen an eine Suchmaschine
- Performance von Oracle Text Funktionen
- Volltextsuche und Navigation im Online Katalog
- Wie geht es noch schneller?



Anforderungen an eine Suchmaschine

- ▶ Performance, Performance, Performance
- ▶ Qualität der Suchergebnisse
 - Sortierung nach Relevanz
 - Häufigkeit der Suchbegriffe
 - genaue Treffer >> unscharfe Treffer
 - Aktionsartikel >> Hauptartikel >> Zubehör
 - Sortierung nach Datum
- ▶ Einfache Bedienung
 - Navigation, Suchfeld, Filter



Agenda

- Anforderungen an eine Suchmaschine
- **Performance von Oracle Text Funktionen**
- Volltextsuche und Navigation im Online Katalog
- Wie geht es noch schneller?



- Ziel: Möglichst wenig Arbeit beim Suchvorgang
 - Maximale Suchdauer festlegen
 - Durchschnittliche Suchdauer festlegen
- Maßnahmen
 - Aufwand auf den Index-/Crawl-Zeitpunkt verlegen
 - bereits verrichteter Aufwand im Cache halten zur Wiederverwendung
 - Datenmenge wenn möglich klein halten



Testtabelle

Scratch Results

#	NR	STADT	BESCHREIBUNG
1	1	Hamburg	Hamburg hat einen großen Hafen
2	2	Berlin	Deutschlands Hauptstadt heißt Berlin
▶ 3	3	Frankfurt/Main	Es gibt Frankfurt am Main und Frankfurt an der Oder
4	4	Köln	Köln hat einen großen Hauptbahnhof
5	5	Ulm	In Ulm und um Ulm und um Ulm herum



Performance: alternate_spelling



- ▶ `ctx_ddl.create_preference('TEST_LEX', 'BASIC_LEXER');`
- ▶ `ctx_ddl.set_attribute('TEST_LEX', 'ALTERNATE_SPELLING', 'GERMAN');`
- ▶ `CREATE INDEX STAEDTE_IDX ON STAEDTE (BESCHREIBUNG) INDEXTYPE IS CTXSYS.CONTEXT PARAMETERS ('LEXER TEST_LEX')`
- ▶ `SELECT * FROM STAEDTE WHERE CONTAINS (BESCHREIBUNG, 'Köln') > 0`
- ▶ `SELECT * FROM STAEDTE WHERE CONTAINS (BESCHREIBUNG, 'koeln') > 0`



Performance: alternate_spelling

#	TOKEN_TEXT	TOKEN_TYPE	TOKEN_FIRST	TOKEN_LAST	TOKEN_COUNT	TOKEN_INFO_LENGTH
2	BERLIN	0	2	2	1	3
3	DER	0	3	3	1	3
4	DEUTSCHLANDS	0	2	2	1	3
5	EINEN	0	1	4	2	6
6	ES	0	3	3	1	3
7	FRANKFURT	0	3	3	1	4
8	GIBT	0	3	3	1	3
9	GROßEN	0	1	4	2	6
10	GROSSEN	0	1	4	2	6
11	HAFEN	0	1	1	1	3

Performance: soundex Operator

- ▶ select * from STAEDTE WHERE CONTAINS (BESCHREIBUNG, '!Deutschland') > 0
- ▶ CTX_QUERY.EXPLAIN

#	EXPLAIN_ID	ID	PARENT_ID	OPERATION	OPTIONS	OBJECT_NAME	POSITION	CARDINALITY
▶ 1	Test	1	0	EQUIVALENCE	(!)	DEUTSCHLAND	1	
2	Test	2	1	WORD		DEUTSCHLAND	1	
3	Test	3	1	WORD		DEUTSCHLANDS	2	



Performance

- Base-Letter Conversion
- Wildcards
 - Query Expansion
 - Prefix- und/oder Substring-Index
- Fuzzy
 - Query Expansion
 - `select * from STAEDTE WHERE CONTAINS (BESCHREIBUNG, 'fuzzy(hauptstadt,,weight)', 1) > 0`



Performance: stem Operator

- Query Expansion
- `ctx_ddl.set_attribute('TEST_LEX', 'INDEX_STEMS', 'GERMAN')`

The screenshot shows a window titled "Scratch Results" containing a table with the following data:

#	TOKEN_TEXT	TOKEN_TYPE	TOKEN_FIRST	TOKEN_LAST	TOKEN_COUNT	TOKEN_INFO_LENGTH
1	DEUTSCHLAND	9	2	2	1	3
2	DIE	9	3	3	1	3
3	EIN	9	1	4	2	6
4	EINE	9	1	4	2	6
5	EINER	9	1	4	2	6
6	GEBEN	9	3	3	1	3
7	GROß	9	1	4	2	6
8	GRÖßE	9	1	4	2	6
9	HABEN	9	1	4	2	6



Performance: stem Operator



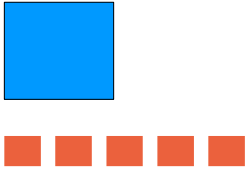
- ▶ select * from STAEDTE WHERE CONTAINS (BESCHREIBUNG, '\$groß', 1) > 0
- ▶ Nur BASIC_LEXER

#	EXPLAIN_ID	ID	PARENT_ID	OPERATION	OPTIONS	OBJECT_NAME	POSITION	CARDINALITY
▶ 1	Test	1	0	EQUIVALENCE	(\$)	GROSS	1	
2	Test	2	1	WORD		GROSS	1	
3	Test	3	1	WORD		GROSSED	2	
4	Test	4	1	WORD		GROSSER	3	
5	Test	5	1	WORD		GROSSES	4	
6	Test	6	1	WORD		GROSSEST	5	
7	Test	7	1	WORD		GROSSING	6	



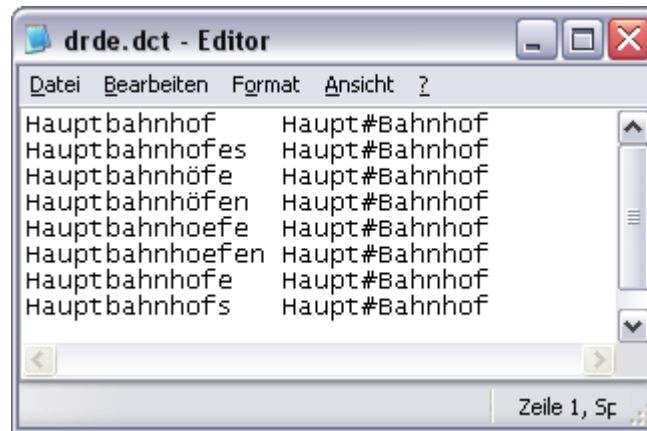
Performance: composite

- `ctx_ddl.set_attribute('TEST_LEX', 'INDEX_STEMS', 'GERMAN');`
- `ctx_ddl.set_attribute('TEST_LEX', 'COMPOSITE', 'GERMAN');`
- `ctx_ddl.create_preference('TEST_WORDLIST', 'BASIC_WORDLIST');`
- `ctx_ddl.set_attribute('TEST_WORDLIST', 'STEMMER', 'GERMAN');`
- Query Expansion
 - Aber Normalisierung der Stemmerinformation in `drde.dct`



Performance: composite

- CONTAINS (beschreibung, 'hauptbahnhof') > 0 (ja)
- CONTAINS (beschreibung, '\$hauptbahnhof') > 0 (ja)
- CONTAINS (beschreibung, 'haupt-bahnhof') > 0 (nein)
- CONTAINS (beschreibung, '\$haupt \$bahnhof') > 0 (nein)
- CONTAINS (beschreibung, '\$haupt AND \$bahnhof') > 0 (ja)
- CONTAINS (beschreibung, '\$bahnhof') > 0 (ja)





Performance: Thesaurus

▶ Query Expansion

- BT
- SYN
- ...



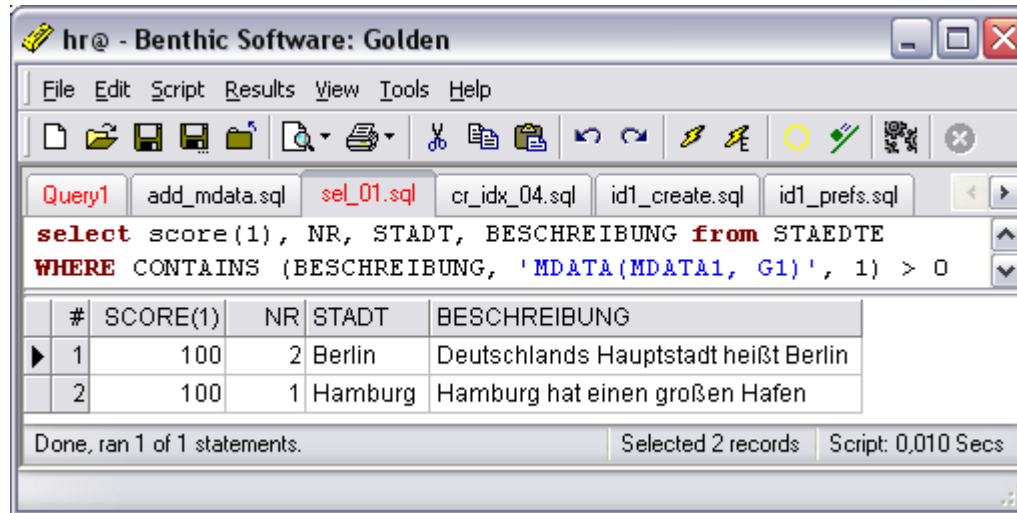
Performance: MDATA

- ◆ Echtes Performance Feature
- ◆ Benötigt keinen `ctx_ddl.sync_index`
- ◆ `TOKEN_TYPE` 400 bis 499 (-400 bis -499)
- ◆ `TOKEN_INFO` sehr kompakt (nur DOC-IDs)
- ◆ Super geeignet für Filter, MDATA = Metadata
- ◆ Mixed Queries, strukturierte Abfragen
- ◆ Ersetzt Field-Section-Filter-Technik von 9i



Performance: MDATA

- `ctx_ddl.create_section_group('TEST_SGP','BASIC_SECTION_GROUP');`
- `ctx_ddl.add_mdata_section('TEST_SGP','MDATA1', 'MDATA1');`
- `ctx_ddl.add_mdata('STAEDTE_IDX', 'MDATA1', 'G1', 'AAAHkhAAEAAAYpHAAA');`
- `ctx_ddl.add_mdata('STAEDTE_IDX', 'MDATA1', 'G2', 'AAAHkhAAEAAAYpHAAA');`





➤ AND Verknüpfung

The screenshot shows a SQL query execution window titled "hr@ - Benthic Software: Golden". The query is as follows:

```
select score(1), NR, STADT, BESCHREIBUNG from STAEDTE
WHERE CONTAINS (BESCHREIBUNG, 'und AND der', 1) > 0
order by score(1) desc
```

The results are displayed in a table with the following columns: #, SCORE(1), NR, STADT, and BESCHREIBUNG. Two records are shown:

#	SCORE(1)	NR	STADT	BESCHREIBUNG
1	4	5	Ulm	In Ulm und um Ulm und um Ulm herum. Ulm liegt an der Donau.
2	4	3	Frankfurt/Main	Es gibt Frankfurt am Main und Frankfurt an der Oder

At the bottom of the window, it states: "Done, ran 1 of 1 statements. Selected 2 records Script: 0,040 Secs".



ACCUMULATE Operator

- ACCUM oder ',' entspricht einem gewichteten ODER

The screenshot shows a database query tool window titled "hr@ - Benthic Software: Golden". The query editor contains the following SQL code:

```
select score(1), NR, STADT, BESCHREIBUNG from STAEDTE
WHERE CONTAINS (BESCHREIBUNG, 'und ,| der', 1) > 0
order by score(1) desc
```

The results are displayed in a table with the following columns: #, SCORE(1), NR, STADT, and BESCHREIBUNG. The table contains three rows of data:

#	SCORE(1)	NR	STADT	BESCHREIBUNG
1	53	5	Ulm	In Ulm und um Ulm und um Ulm herum. Ulm liegt an der Donau.
2	52	3	Frankfurt/Main	Es gibt Frankfurt am Main und Frankfurt an der Oder
3	2	4	Köln	Köln hat einen großen Hauptbahnhof der neben dem Dom ist.

At the bottom of the window, it shows "Done, ran 1 of 1 statements." and "Selected 3 records | Script: 0,030 Secs".



➤ Kombination: Volltextrecherche mit Filter

The screenshot shows the Golden Software interface with a SQL query window. The query is as follows:

```
select score(1), NR, STADT, BESCHREIBUNG from STAEDTE
WHERE CONTAINS (BESCHREIBUNG, 'und , der', 1) > 0
AND CONTAINS (BESCHREIBUNG, 'und AND der') > 0
order by score(1) desc
```

The results table shows two records:

#	SCORE(1)	NR	STADT	BESCHREIBUNG
1	53	5	Ulm	In Ulm und um Ulm und um Ulm herum. Ulm liegt an der Donau.
2	52	3	Frankfurt/Main	Es gibt Frankfurt am Main und Frankfurt an der Oder

Done, ran 1 of 1 statements. Selected 2 records Script: 0,150 Secs



➤ Kombination: Volltextrecherche mit Filter

- Sort funktioniert nur bei einem Begriff im Suchfeld!

The screenshot shows a database query tool window titled "hr@ - Benthic Software: Golden". The query executed is:

```
select score(1), NR, STADT, BESCHREIBUNG from STAEDTE
WHERE CONTAINS (BESCHREIBUNG, 'der AND MDATA(MDATA1, G2)', 1) > 0
order by score(1) desc
```

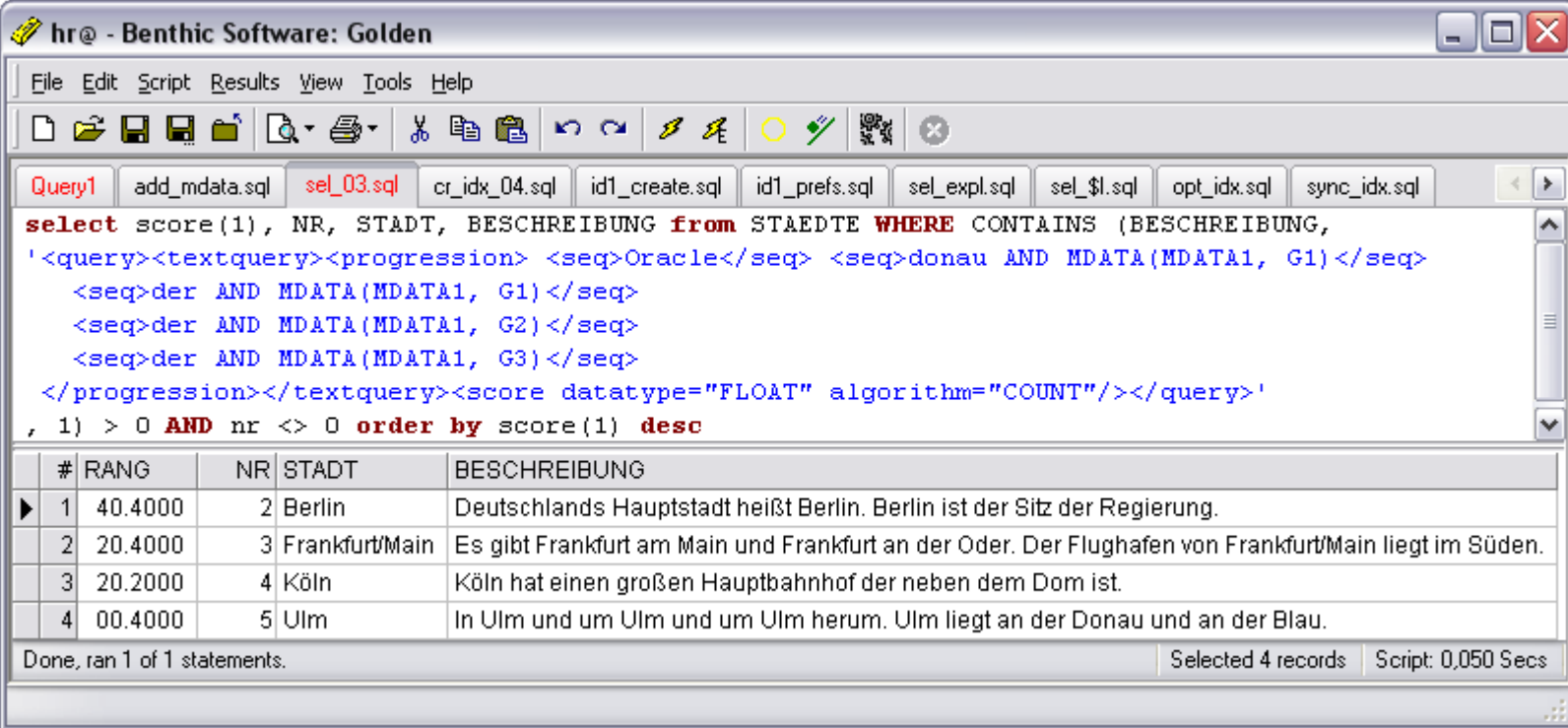
The results are displayed in a table with the following columns: #, SCORE(1), NR, STADT, and BESCHREIBUNG. The table contains three rows of data:

#	SCORE(1)	NR	STADT	BESCHREIBUNG
1	7	2	Berlin	Deutschlands Hauptstadt heißt Berlin. Berlin ist der Sitz der Regierung.
2	7	3	Frankfurt/Main	Es gibt Frankfurt am Main und Frankfurt an der Oder. Der Flughafen von Frankfurt/Main liegt im Süden.
3	3	4	Köln	Köln hat einen großen Hauptbahnhof der neben dem Dom ist.

The status bar at the bottom indicates "Done, ran 1 of 1 statements." and "Selected 3 records | Script: 0,020 Secs".

Query Relaxation

- Echtes Performance Feature
- Bug: Keine Progression, falls 1. Set keine Treffer liefert



The screenshot shows the GoldenGate software interface. The title bar reads "hr@ - Benthic Software: Golden". The menu bar includes "File", "Edit", "Script", "Results", "View", "Tools", and "Help". The toolbar contains various icons for file operations and execution. The main window displays a query editor with the following SQL code:

```
select score(1), NR, STADT, BESCHREIBUNG from STAEDTE WHERE CONTAINS (BESCHREIBUNG,
'<query><textquery><progression> <seq>Oracle</seq> <seq>donau AND MDATA(MDATA1, G1)</seq>
<seq>der AND MDATA(MDATA1, G1)</seq>
<seq>der AND MDATA(MDATA1, G2)</seq>
<seq>der AND MDATA(MDATA1, G3)</seq>
</progression></textquery><score datatype="FLOAT" algorithm="COUNT"/></query>'
, 1) > 0 AND nr <> 0 order by score(1) desc
```

Below the query editor, a table displays the results of the query:

#	RANG	NR	STADT	BESCHREIBUNG
1	40.4000	2	Berlin	Deutschlands Hauptstadt heißt Berlin. Berlin ist der Sitz der Regierung.
2	20.4000	3	Frankfurt/Main	Es gibt Frankfurt am Main und Frankfurt an der Oder. Der Flughafen von Frankfurt/Main liegt im Süden.
3	20.2000	4	Köln	Köln hat einen großen Hauptbahnhof der neben dem Dom ist.
4	00.4000	5	Ulm	In Ulm und um Ulm und um Ulm herum. Ulm liegt an der Donau und an der Blau.

At the bottom of the window, a status bar indicates "Done, ran 1 of 1 statements." and "Selected 4 records Script: 0,050 Secs".



Agenda

- Anforderungen an eine Suchmaschine
- Performance von Oracle Text Funktionen
- **Volltextsuche und Navigation im Online Katalog**
- Wie geht es noch schneller?

➤ Reduzierung von Information

- Keine Unterscheidung von Groß-/Kleinschreibung
- Konvertierung von Umlauten und diakritischen Zeichen
- Wortsegmentierung mit Stoppwort-Elimination

➤ Normalisierung von Information

- Stemmer, Kompositazerlegung, phon. Algorithmus

➤ Anreicherung von Informationen

- Synonyme, Katalogknoten, Auflösung v. Abkürzungen

➤ Tabellen

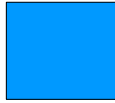
- Wortliste, Synonymliste, Wortzerlegungen, Stopwortliste, Zensurliste, Menü-Hierarchie
- Attributklassen, Attributwerte, Zuordnung zum Menü
- Artikel-Kopf-Tabelle (ähnlich wie Materialized View)
- Suchergebnisse (Tabellen bzw. Pipelined Table Functions)
 - Query Protokoll
 - Treffer Tabelle
 - Trefferverteilung auf Attribute



Suchfunktionen Online Katalog



- ◆ Direkte Suche (Volltextrecherche)
 - Suchbegriffe im Suchfeld
- ◆ „Stöbern“ (Strukturierte Suche / Filter setzen)
 - Menü-Bäume
 - Auswahlassistent
- ◆ Kombination



Demo Online Katalog

The screenshot shows a Mozilla Firefox browser window with the title 'Online Katalog - Mozilla Firefox'. The address bar contains the URL 'http://127.0.0.1:8080/ci/ci.w'. The main content area displays the 'Doag Testsystem' interface. On the right side, there is a blue button labeled 'Online Katalog'. Below this, there is a search section with a text input field and a 'Suchen' button. To the left of the search field is a table with a header 'Auswahl' and four rows of product categories. At the bottom left, there is a copyright notice '© 2007 ci database gmbh' and a 'Fertig' button. At the bottom right, there is a 'Seitenanfang' link.

Auswahl
Möbelbeschläge, Möbelgriffe
Baubeschläge
Montagetechnik
Elektronisches Schließsystem Dialock

Suche:

Bitte geben Sie Ihre Suchbegriffe ein.

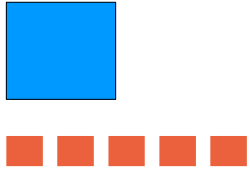
© 2007 ci database gmbh [Seitenanfang](#)

Fertig



Agenda

- Anforderungen an eine Suchmaschine
- Performance von Oracle Text Funktionen
- Volltextsuche und Navigation im Online Katalog
- **Wie geht es noch schneller?**



Wie geht es noch schneller?

- Function-Based Bitmap Indexes?
- neue Domain Indexes?
- Relevanzoptimierung
 - bei AND Verknüpfung
 - im Bereich Query Relaxation