

Oracle Text

Katerina Mpalaska

Senior Sales Consultant

Server Technologies Competence Center Stuttgart

Oracle Deutschland GmbH

ORACLE®

Neues in Oracle Text 10g

Oracle TEXT

Dokumente hochladen Suche

Suche nach:

Suchen

Ergebnisse

-  **100** [HTML DB für die "Finanz"](#) 14.06.2005 application/msword
[skalierbar. An dieser Stelle setzt die Oracle HTML DB an. Mit der HTML DB können datenbankgestützte Anwendungen](#)
-  **57** [5 Gründe für die HTML DB \(Fach\)](#) 14.06.2005 application/msword
[5Gründe warum Sie Oracle HTML DB anstelle von Access einsetzen sollten...warum Sie ORACLE HTML DB anstelle von Access](#)
-  **50** [5 Gründe für die HTML DB \(IT\)](#) 14.06.2005 application/msword
[5Gründe warum Sie Oracle HTML DB anstelle von Access einsetzen sollten...warum Sie ORACLE HTML DB anstelle von Access](#)
-  **7** [HTML DB Schnellanwendung](#) 14.06.2005 application/msword
[stellt sie alle wesentlichen Komponenten der HTML DB dar. Labor-1](#)

1 - 4

ORACLE

Neues in Oracle Text 10g

Index erstellen

Volltextrecherche

Monitoring /Wartung

Neues in Oracle Text 10g

Index erstellen

Volltextrecherche

Monitoring / Wartung

Oracle Text 10g – 10.1.04, 10.2.01

Neuer AUTO-Filter

ersetzt den INSO-Filter

```
create index hdocsx on hdocs(text) indextype is
  ctxsys.context
parameters ('datastore ctxsys.file_datastore
  filter ctxsys.auto_filter
  format column binary);
```

- INSO-Filter Funktionalität wird mit AUTO-Filter weiterhin unterstützt
- Für neue Applikationen: AUTO-Filter nutzen !
- Formate: s. Oracle Text Reference Guide
http://download.oracle.com/docs/cd/B19306_01/text.102/b14218.pdf

Beispiel: Auto-Filter

Oracle Text 10gR1

Unicode Lexer

- Neue Lexer Präferenz – **World_Lexer**
- Support für jede Unicode 4.0 Sprache



Oracle Text 10gR1 World Lexer

Beispiel

```
INSERT INTO globaldoc2 VALUES(1,'Dies ist ein deutsches Dokument');
INSERT INTO globaldoc2 VALUES(2,'This is an english document');

exec ctx_ddl.create_preference('MYLEXER', 'world_lexer');
create index globaldoc2_idx on globaldoc2 (text) indextype is ctxsys.CONTEXT
  parameters ('lexer MYLEXER stoplist CTXSYS.EMPTY_STOPLIST');

SELECT * FROM globaldoc2 WHERE contains(text, 'document')>0;
```

Ergebnis:

```
  doc_id text
-----
      2 This is an english document
```

Beispiel: World-Lexer

Oracle Text 10gR1

Erweiterungen - Mail Filtering

- Konvertieren einer RFC-2045, RFC-822 Email in ein indizierbares Format
- Behandlung der Message Bodies basierend auf dem Content-Type
 - Text Meldungen werden in den DB Characterset konvertiert
 - Binärer Text wird gefiltert

```
CTX_DDL.CREATE_PREFERENCE('my_mail_filt', 'mail_filter');  
CTX_DDL.SET_ATTRIBUTE ('my_mail_filt', 'INDEX_FILES', 'subject');  
CTX_DDL.SET_ATTRIBUTE ('my_mail_filt', 'PART_FIELD_STYLE', 'ignore');
```



Oracle Text 10gR1

Erweiterungen - Mail Filtering

- Mail File Configuration File Structure (drmailfl.txt)
unter ORACLE_HOME/ctx/config

[behavior]

- application/zip IGNORE
- application/msword AUTO_FILTER
- ...

[charsets]

- US-ASCII US7ASCII
- ISO-8859-1 WE8ISO8859P1
- ...



Oracle Text 10gR1 Mail Filtering

To: somebody@someplace
Subject: mainheader
Content-Type: multipart/mixed
...

...
Content-Type: text/plain
X-Ref: some_value
Subject: subheader 1
...

...
Content-Type: text/plain
X-Control: blah blah blah
Subject: subheader 2

```
PART_FIELD_STYLE => IGNORE
  <SUBJECT>mainheader</SUBJECT>

PART_FIELD_STYLE => TAG
  <SUBJECT>mainheader</SUBJECT>
  <SUBJECT>subheader1</SUBJECT>
  <SUBJECT>subheader2</SUBJECT>

PART_FIELD_STYLE => FIELD
  <SUBJECT>mainheader<SUBJECT>
  SUBJECT:subheader1
  SUBJECT:subheader2

PART_FIELD_STYLE => TEXT
  <SUBJECT>mainheader</SUBJECT>
  subheader1
  subheader2
```



Oracle Text 10gR1

MDATA

- Abfrage von Metadaten innerhalb der Text-Abfrage
- Neuer Section-Typ: MDATA

MDATA(sectionname, value)

```
SELECT book_info from library_stock
where contains (book_info,
'Irving within author and mdata(status, In Stock)') > 0;
```

Oracle Text 10gR2

- Neue Stopwörter für English, Finnisch, Italiensich, Spanisch und Schwedisch
- CTXRULE-Unterstützung für Chinesisch, Japanisch, Koreanisch
- Japanische Sprachunterstützung
 - Delimiter characters
 - Fuzzy Function fürs Japanische
 - Japanisches Stemming
 - Japanischer Unicode
 - Japanisches Benutzer Lexikon

Neues in Oracle Text 10g

Index erstellen

Volltextrecherche

Monitoring / Wartung

Oracle Text 10gR1

Neue Deutsche Rechtschreibung

- Alte Schreibweise
- Neue Schreibweise

Alte Rechtschreibung	Neue Rechtschreibung
Potential	Potenzial
Katarrh	Katarr, Katarrh
Delphin	Delfin
Erdgeschoß	Erdgeschoss
Schiffahrt	Schifffahrt
Weh tun	Wehtun
Irgend etwas	Irgendetwas
Soviel	So viel

Oracle Text 10gR1

Neue Deutsche Rechtschreibung

Beispiel

```
INSERT INTO Tabled VALUES (1, 'Känguruh daß Potential Schifffahrt');

exec ctx_ddl.create_preference('ngslx','basic_lexer');
exec ctx_ddl.set_attribute('ngslx','new_german_spelling','true');

CREATE INDEX ngslx_idx ON Tabled (document) INDEXTYPE IS CTXSYS.CONTEXT
PARAMETERS ('lexer ngslx');

SELECT * FROM Tabled WHERE contains(document, 'Potenzial')>0;
```

Ergebnis:

item_id	document
1	Känguruh daß Potential Schifffahrt

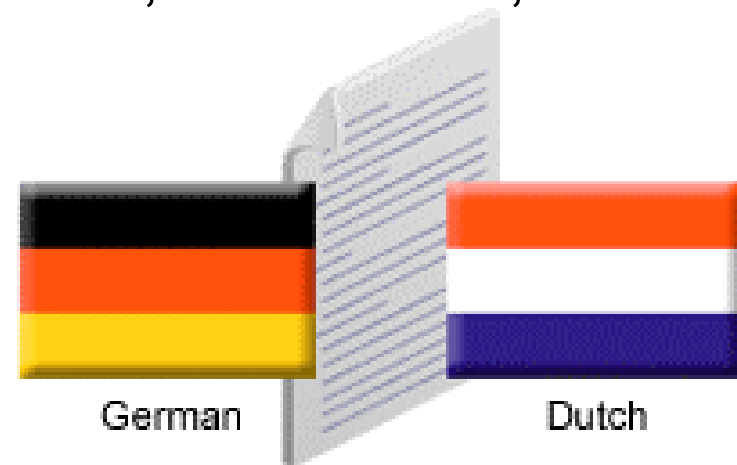


Beispiel: Neue Rechtschreibreform

Oracle Text 10gR1

Komposita - Unterstützung

- Suche nach zusammengesetzten Wörtern
- Für Deutsch und Holländisch
- Attribut der BASIC_LEXER Präferenz: ***composite***
- Suche nach Bahnhof:
Ergebnis: Hauptbahnhof, Nordbahnhof, Ostbahnhof,...





Beispiel: Composite

Oracle Text 10gR2

Text-Fragmente – neue Prozeduren:

- Key Word in Context (KWIC)
- CTX_DOC.SNIPPET, CTX_DOC.Policy_SNIPPET
- Schlüsselwörter und umliegender Text
- Unterschiede zu MARKUP oder HIGHLIGHT

Oracle Text 10gR2

```
CTX_DOC.SNIPPET (  
    index_name    IN VARCHAR2 ,  
    textkey       IN VARCHAR2 ,  
    text_query    IN VARCHAR2 ,  
    starttag      IN VARCHAR2 DEFAULT '<b>',  
    endtag        IN VARCHAR2 DEFAULT '</b>',  
    entity_translation  
                    IN BOOLEAN DEFAULT TRUE ,  
    separator     IN VARCHAR2 DEFAULT '<b>...</b>'  
)  
return varchar2;
```



Beispiel: SNIPPET

Oracle Text 10gR1

Query Template - Erweiterungen

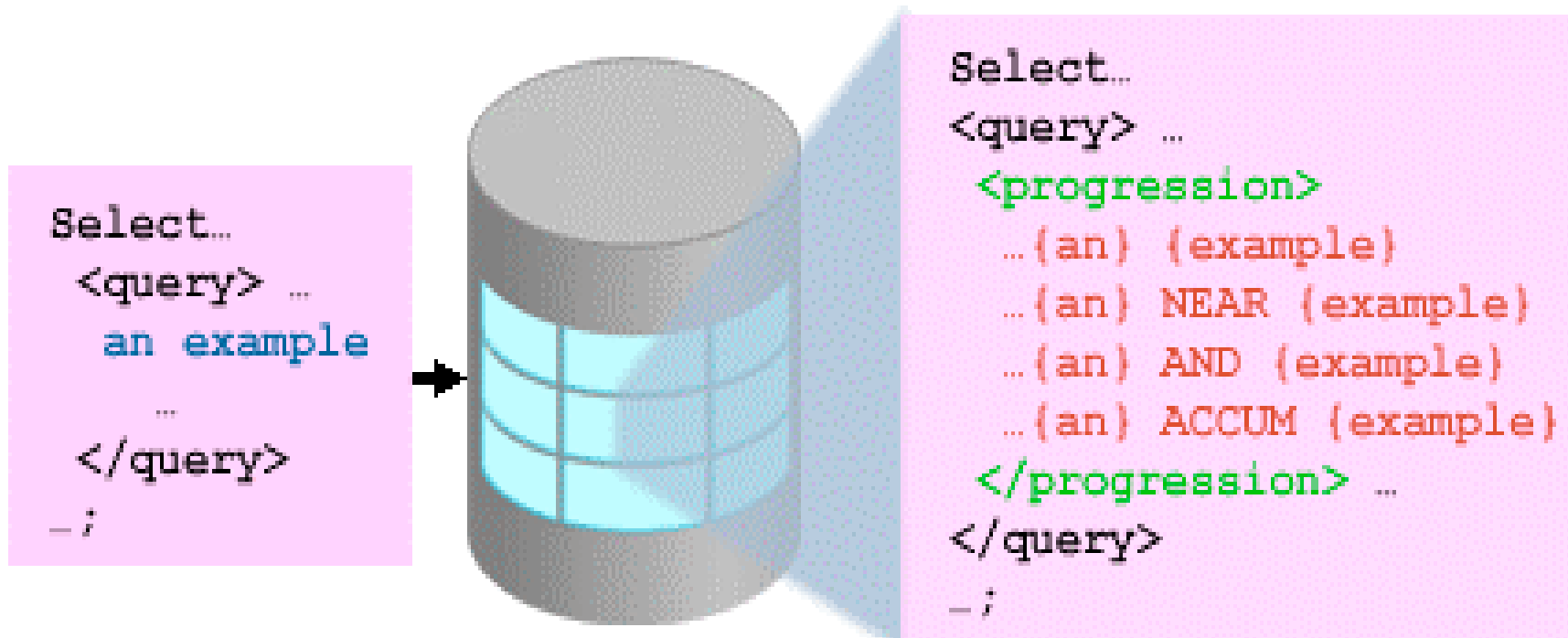
- Query Rewrite
- Progressive Relaxation
- Spezifikation - Query Language
- Alternatives Scoring
- Alternative Grammatik



`<query>... .</query>`

Oracle Text 10gR1

Progressive Relaxation



Oracle Text 10gR1

Progressive Relaxation

```
select * from purchaseorders where CONTAINS
(text, '<query>
  <textquery lang="ENGLISH" grammar="CONTEXT">
    Retail Sales
    <progression>
      <seq>{Retail} {Sales}</seq>
      <seq>{Retail} NEAR {Sales}</seq>
      <seq>{Retail} AND {Sales}</seq>
      <seq>{Retail} ACCUM {Sales}</seq>
    </progression>
  </textquery>
  <score datatype="INTEGER" algorithm="COUNT"/>
</query>')>0;
```



Beispiel: Progressive Relaxation

Oracle Text 10gR1

Query Templates - Erweiterungen

- **Multi_Lexer - Query Language**

```
select id from docs where CONTAINS (text,  
'<query><textquery lang="french">bon soir</textquery></query>')>0;
```

- **Alternatives Scoring**

```
select id from docs where CONTAINS (text,  
'<query>  
<textquery grammar="CONTEXT" lang="english"> mustang </textquery>  
<score datatype="float" algorithm="DEFAULT"/>  
</query>')>0
```

- **Alternative Grammatik**

```
select id from docs where CONTAINS (text,  
'<query>  
<textquery grammar="CTXCAT">San Diego</textquery>  
<score datatype="integer"/>  
</query>')>0;
```

Oracle Text 10gR1

INPATH and HASPATH Erweiterungen

- Highlighting mit
 - INPATH
 - HASPATH



Neues in Oracle Text 10g

Index erstellen

Volltextrecherche

Monitoring /Wartung

Oracle Text 10gR2

- **Alter Index - neue Parameter**

ALTER INDEX PARAMETERS

(Änderung der Parameter ohne Wiederaufbau des Index) und

ALTER INDEX MODIFY PARTITION PARAMETERS

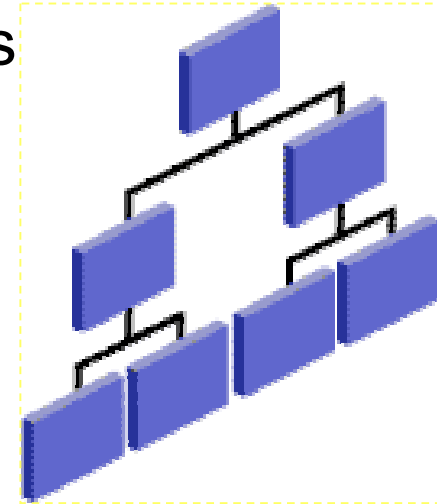
- **CTX_ADM.MARK_FAILED**

Index Status von Loading auf Failed gesetzt, um Recovery mit Resume_Index zu ermöglichen

Oracle Text 10gR1

Alter Index Rebuild Replace Metadata

- Ersetzen der existierenden Präferenz-Settings durch neue Präferenz-Settings
- betrifft auch die SYNC Parameter
- Kein Neuaufbau des Textindexes

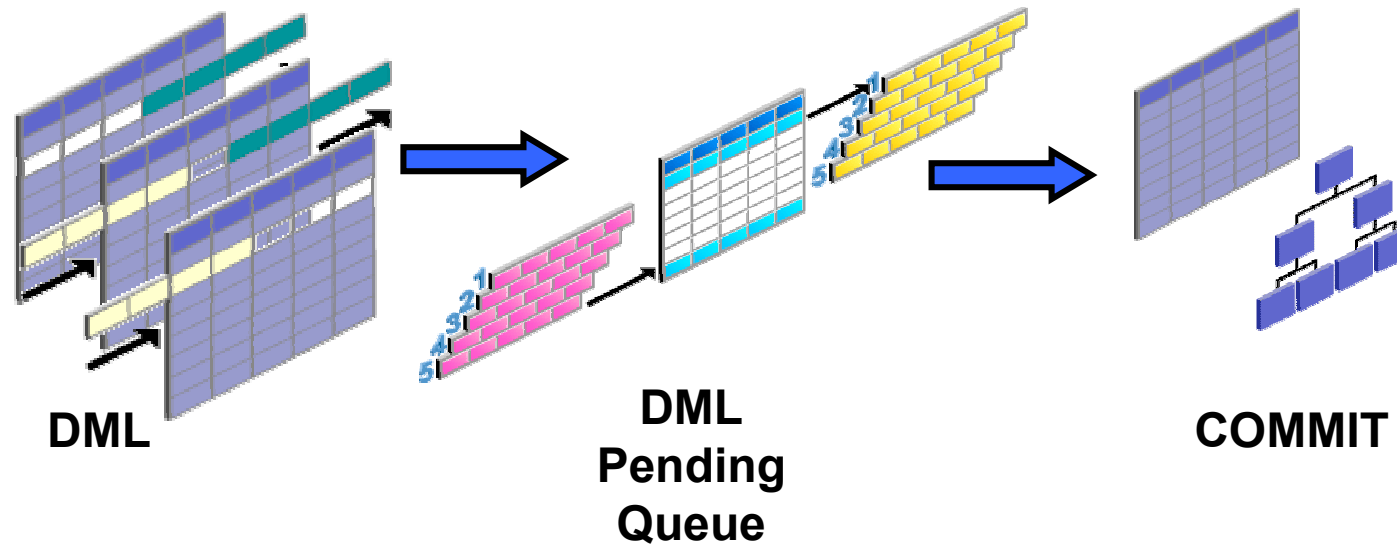


```
ALTER INDEX text_idx  
REBUILD PARAMETERS ('replace metadata language  
column lang lexer m_lexer');
```

ACHTUNG – eigene Verantwortung für einen konsistenten Index

Oracle Text 10gR1

Auto und ON COMMIT Synchronisierung



- SYNC (MANUAL | EVERY "*interval-string*" | ON COMMIT)

Oracle Text 10gR1

Änderung der Rechte für CTXSYS

~~DBA Privilege~~



Oracle Text 10gR1

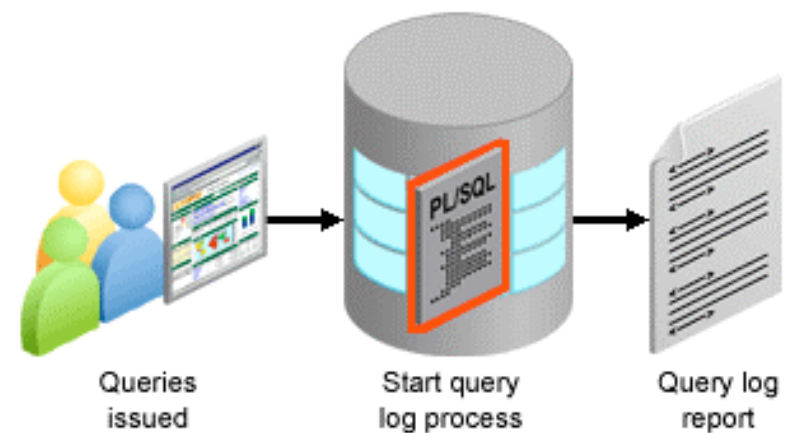
CTX_REPORT Package

- CTX_REPORT.DESCRIBE_INDEX
- CTX_REPORT.DESCRIBE_POLICY
- CTX_REPORT.CREATE_INDEX_SCRIPT
- CTX_REPORT.CREATE_POLICY_SCRIPT
- CTX_REPORT.INDEX_SIZE
- CTX_REPORT.INDEX_STATS
- CTX_REPORT.TOKEN_INFO
- CTX_REPORT.QUERY_LOG_SUMMARY
- CTX_REPORT.TOKEN_TYPE

Oracle Text 10gR1

CTX_REPORT - Query Log Analyse

- Welche Abfragen wurden gemacht ?
- Welche Abfragen waren erfolgreich ?
- Welche Abfragen waren nicht erfolgreich ?
- WAS wurde WIE HÄUFIG angefragt ?

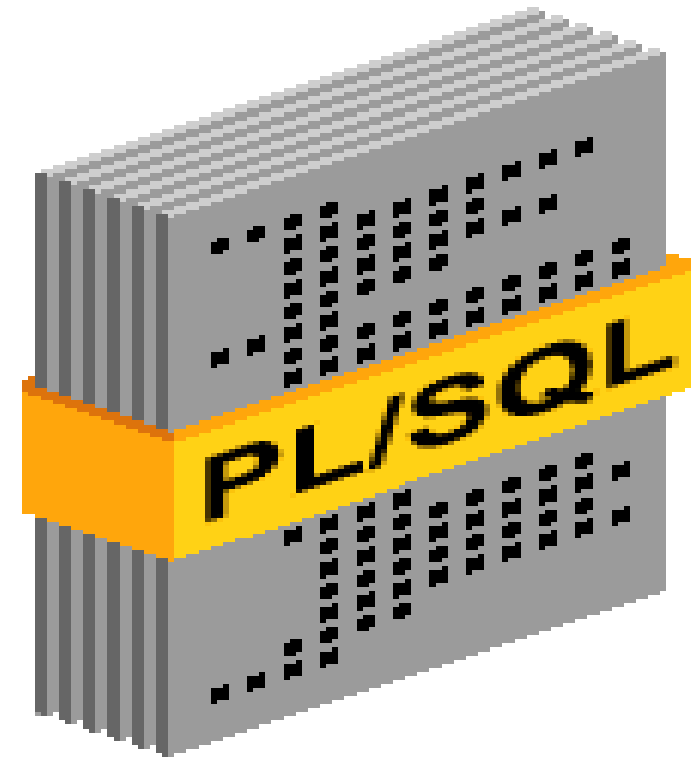


Oracle Text 10gR1

CTX_REPORT - Query Analyse

- Start query logging
- Abfragen starten
- End query logging
- Query log summary

```
procedure CTX_REPORT.QUERY_LOG_SUMMARY(  
  logfile          IN VARCHAR2,  
  indexname        IN VARCHAR2 DEFAULT NULL,  
  result_table     IN OUT NOCOPY QUERY_TABLE,  
  row_num          IN NUMBER,  
  most_freq        IN BOOLEAN DEFAULT TRUE,  
  has_hit          IN BOOLEAN DEFAULT TRUE  
);
```



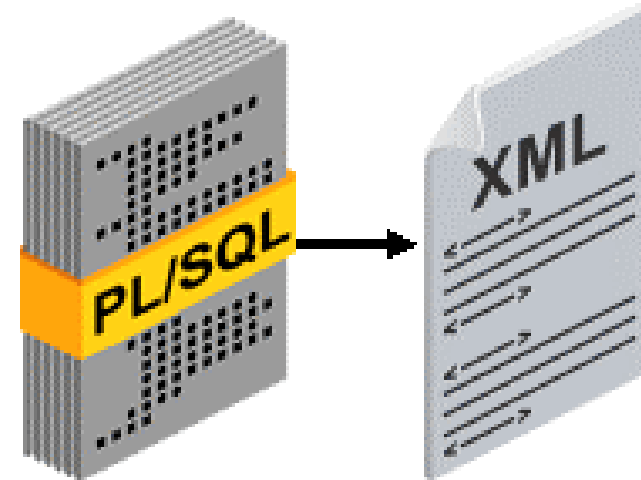
CTX_REPORT

Oracle Text 10gR1

XML Output - CTX_REPORT Package

```
SELECT ctx_report.describe_index('DOCS_TEXT','XML') FROM dual;  
CTX_REPORT.DESCRIBE_INDEX('DOCS_TEXT','XML')
```

```
<CTXREPORT>  
<DESCRIBE_INDEX>  
<INDEX_ATTRIBUTES>  
<INDEX_ATTRIBUTE NAME="index name">  
"CTXSYS"."DOCS_TEXT"  
</INDEX_ATTRIBUTE>  
<INDEX_ATTRIBUTE NAME="index id">  
1392  
</INDEX_ATTRIBUTE>  
<INDEX_ATTRIBUTE NAME="index type">  
context  
...
```



CTX_REPORT

Beispiele: CTX_Report

Neues in Oracle Text 10g

Index erstellen

Volltextrecherche

Monitoring /Wartung

F r a g e n ?

Weitere Informationen zu Oracle Text unter

<http://www.oracle.com/technology/products/text/index.html>

Email:

katerina.mpalaska@oracle.com