



**ORACLE®**

## **Oracle Text 概要**

日本オラクル株式会社

**Oracle** Direct

# 検索機能 ～ テキスト検索と部分一致検索の違い

## ➤ 部分一致検索

- ✓ パターン・マッチング全角・半角・大文字・小文字の区別あり
- ✓ CHAR/VARCHAR2型のみ検索対象となる

## ➤ テキスト検索

- ✓ 予め検索対象テキストから作成しておいたテキスト索引を使用し検索
- ✓ 基本的に、全角・半角・大文字・小文字の区別なし
- ✓ CHAR/VARCHAR2/CLOB/BLOB/BFILE型などが検索対象となる

### 部分一致検索

```
SELECT * FROM doc  
WHERE text LIKE '%菅直人%';
```

文書2のみヒット

### 本文

1: ...菅 直人さんは選挙で、...  
2: ...挨拶で、菅直人さんは...

### テキスト検索

```
SELECT * FROM doc  
WHERE CONTAINS(text,'菅直人') > 0;
```

文書1, 2共にヒット

スペース、「・」、「/」など記号の有無を無視した検索が可能

# 検索機能 ～ スコア表示、ヒット順表示

## ➤ スコア

- ✓ ヒットした文書がどの程度検索語句を含んでいるかを表す整数値 (0～100)
  - 大きい値ほど条件に近い
  - 結果の比較などに使用し、相対的な値
  - 指定した語を含むテキストは必ず0より大きなスコアが付く
  - CONTAINS 演算子の戻り値で表示  
→ CONTAINS(...) > 0 と指定する

## ➤ ヒット順表示

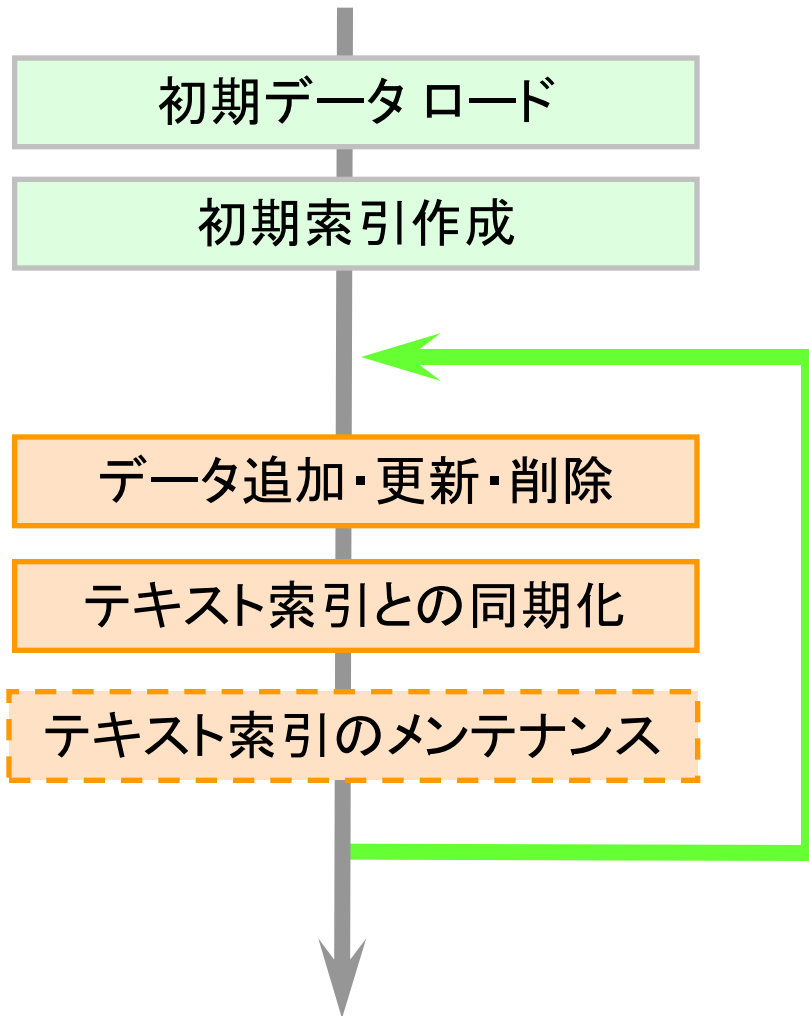
- ✓ スコア順にソートして、結果表示
- ✓ スコアの上位n件を表示可能

### スコア表示

```
SELECT title,SCORE(1) FROM doc  
WHERE CONTAINS( text, 'java', 1) > 0  
ORDER BY SCORE(1);
```

CONTAINS 演算子の戻り値が "0より大" と指定

# Oracle Text 使用手順 ～ 一連の流れ



- 環境作成
  - ✓ 初期データロード
  - ✓ 初期索引作成
- 索引メンテナンス
  - ✓ データ追加・更新・削除
    - リアルタイム or バッチで行うことが可能
  - ✓ 索引の同期化 (SYNC)
    - 検索表にデータが追加され、新しく抽出されたトークンが、トークン表の末尾に格納
    - リアルタイム or バッチで行うことが可能
  - ✓ テキスト索引のメンテナンス (OPTIMIZE)
    - フラグメンテーションを解消し、トークン表が最適な状態に保つ
    - 最適化処理 or 索引の再作成で行うことが可能

# Oracle Text 使用手順 ～ 初期データロード

- 検索対象の表を作成 (表名: Master)

ID	Title	Author	Publisher	Publish_date	Abstract
1	.....	.....	.....	.....	.....
2	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
XXXXX	.....	.....	.....	.....	.....

- 全文検索の対象とする列
  - ✓ Title
- 既に以下のものがインストールされているものとする
  - ✓ Oracle Database

□ OptionとしてOracle Textが選択されインストール済み

※ Oracle Textは、データベースインストール後でも、インストール可能

# Oracle Text 使用手順 ～ 初期索引作成 1

1. 利用するユーザに以下の権限を与える
  - CTXAPP ロール

```
SQL> CONNECT / AS SYSDBA
SQL> GRANT CTXAPP TO <ユーザ名>;
```

2. プリファレンス(全文検索対象のデータがどんなものかを示す指定)を作成

```
SQL> CONNECT <ユーザ名>/<パスワード>
SQL> BEGIN
  2   CTX_DDL.CREATE_PREFERENCE(' <プリファレンス名>',
  3                               ' JAPANESE_VGRAM_LEXER' );
  4 END;
  5 /
```

これで「データが日本語です」という指定になる

- ※ 今回の場合は、以下の項目を入力。  
<プリファレンス名>: 任意

# Oracle Text 使用手順 ～ 初期索引作成 2

## 3. 全文検索用の索引を作成

- これによって、本で表すと巻末の索引を作成します。  
全文検索は、この索引を使用して検索が行われます

```
SQL> CREATE INDEX <索引名> ON <表名>(<列名:>)  
2     INDEXTYPE IS CTXSYS.CONTEXT  
3     PARAMETERS (' LEXER <プリファレンス名> ');
```

索引が作成されました。

- ※ 今回の場合は、以下の項目を入力。  
<表名>: Master, <列名>: 書籍、著者名  
<プリファレンス名>: 2.と同じ名前

## 4. 全文検索が実行可能

```
SQL> SELECT <列名> FROM <表名>  
2     WHERE CONTAINS (<列名>, ' <検索文字列> ') > 0;
```

# 索引メンテナンス ～ サンプルデータ

- 検索対象表 (Master) に対し、「Oracle Text 使用手順」と同様の方法で、全文検索用の索引を作成

検索対象表 (表名: Master)

ID	Title	...
1	ORACLEデータベース管理	...
2	簡単ORACLEチューニング	...

トークン表

トークン	出現位置
ORACLE	(1,1), (2,2)
データ	(1,2)
タブ	(1,3)
ベース	(1,4)
管理	(1,5)
簡単	(2,1)
チューニ	(2,3)
ニング	(2,4)

※ (1,1)は、(ID, トークン位置)



# 索引メンテナンス ～ 新規文書情報の追加

## ▶ テキスト索引の同期化

- ✓ リアルタイムでも可能だが、24時間365日の運用を行わない限り  
バッチ処理で行う方がより効果的

検索対象表 (表名: Master)

ID	Title	..
1	ORACLEデータ ベース管理	..
2	簡単ORACLEチ ューニング	..
3	ORACLE バックアップ	·

追加

新たに追加されたトークン  
が索引の末尾に追加

```
EXEC CTX_DDL.SYNC_INDEX( <インデックス名> );
```

トークン表

トークン	出現位置
ORACLE	(1,1), (2,2)
データ	(1,2)
タベ	(1,3)
ベース	(1,4)
管理	(1,5)
簡単	(2,1)
チューニ	(2,3)
ング	(2,4)
ORACLE	(3,1)
バック	(3,2)
クア	(3,3)
アップ	(3,4)

# 索引メンテナンス ~ 新規文書情報の追加

## ▶ テキスト索引のメンテナンス

トークン表

トークン	出現位置
ORACLE	(1,1), (2,2)
データ	(1,2)
タブ	(1,3)
ベース	(1,4)
管理	(1,5)
簡単	(2,1)
チューニ	(2,3)
ニング	(2,4)
ORACLE	(3,1)
バック	(3,2)
クア	(3,3)
アップ	(3,4)

テキスト索引に  
複数エントリ  
されている  
同一トークン  
をまとめる



トークン表

トークン	出現位置
ORACLE	(1,1), (2,2), <b>(3,1)</b>
データ	(1,2)
タブ	(1,3)
ベース	(1,4)
管理	(1,5)
簡単	(2,1)
チューニ	(2,3)
ニング	(2,4)
バック	(3,2)
クア	(3,3)
アップ	(3,4)

```
EXEC CTX_DDL.OPTIMIZE_INDEX(<インデックス名>, 'FAST');
```

# 索引メンテナンス ~ 文書情報の削除

## ▶ テキスト索引の同期化

検索対象表 (表名: Master)

ID	Title	..
1	ORACLEデータ ベース管理	..
2	簡単ORACLEチ ューニング	..
3	ORACLE バックアップ	..

追加

トークン表

トークン	出現位置	
ORACLE	(1,1), (2,2), (3,1)	削除済
データ	(1,2)	
タベ	(1,3)	
ベース	(1,4)	
管理	(1,5)	
簡単	(2,1)	削除済
チューニ	(2,3)	削除済
ニング	(2,4)	削除済
バック	(3,2)	
クア	(3,3)	
アップ	(3,4)	

検索対象表への削除処理を、  
Oracle Text は即座に検知し  
て検索結果に反映させる

※ 文書情報の削除は、リアルタイムに検索結果に反映されるため、コマンドの実行は不要。

# 索引メンテナンス ~ 文書情報の削除

## ▶ テキスト索引のメンテナンス

トークン表

トークン	出現位置
ORACLE	(1,1), <del>(2,2)</del> , (3,1)
データ	(1,2)
タブ	(1,3)
ベース	(1,4)
管理	(1,5)
簡単	<del>(2,1)</del>
チューニ	<del>(2,3)</del>
ニング	<del>(2,4)</del>
バック	(3,2)
クア	(3,3)
アップ	(3,4)

削除されて既に  
検索対象表に  
存在していない  
トークンを削除



トークン表

トークン	出現位置
ORACLE	(1,1), (3,1)
データ	(1,2)
タブ	(1,3)
ベース	(1,4)
管理	(1,5)
バック	(3,2)
クア	(3,3)
アップ	(3,4)

```
EXEC CTX_DDL.OPTIMIZE_INDEX(<インデックス名>, 'FULL');
```

# 参考情報

## ➤ 製品マニュアル

### ✓ 『Oracle Text リファレンス』

[http://otndnld.oracle.co.jp/document/products/oracle11g/111/doc\\_dvd/text.111/E05789-02/toc.htm](http://otndnld.oracle.co.jp/document/products/oracle11g/111/doc_dvd/text.111/E05789-02/toc.htm)

### ✓ 『Oracle Text アプリケーション開発者ガイド』

[http://otndnld.oracle.co.jp/document/products/oracle11g/111/doc\\_dvd/text.111/E05788-02/toc.htm](http://otndnld.oracle.co.jp/document/products/oracle11g/111/doc_dvd/text.111/E05788-02/toc.htm)

## ➤ OTN (Oracle Technology Network)

### ✓ Oracle Text 製品情報

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



以上の事項は、弊社の一般的な製品の方向性に関する概要を説明するものです。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。以下の事項は、マテリアルやコード、機能を提供することをコミットメント(確約)するものではないため、購買決定を行う際の判断材料になさらないで下さい。オラクル製品に関して記載されている機能の開発、リリースおよび時期については、弊社の裁量により決定されます。

OracleとJavaは、Oracle Corporation 及びその子会社、関連会社の米国及びその他の国における登録商標です。文中の社名、商品名等は各社の商標または登録 商標である場合があります。