



PL/PDF Report Indexing

v2.3.1

Oracle Text

Oracle Text uses standard SQL to index, search, and analyze text and documents stored in the Oracle database, in files, and on the web. Oracle Text can perform linguistic analysis on documents, as well as search text using a variety of strategies including keyword searching, context queries, Boolean operations, pattern matching, mixed thematic queries, HTML/XML section searching, and so on. It can render search results in various formats including unformatted text, HTML with term highlighting, and original document format. Oracle Text supports multiple languages and uses advanced relevance-ranking technology to improve search quality. Oracle Text also offers advanced features like classification, clustering, and support for information visualization metaphors.

PL/PDF can store PDF reports in the Oracle database, so it can use Oracle Text for indexing created reports. The examples use "Report Manager" tables in this article.

CONTEXT Indexes

The CONTEXT index type is used to index large amounts of text such as Word, PDF, XML, HTML or plain text documents. In this example we will store the data in a BLOB column, which allows us to store binary documents like PDF as well as plain text.

First we create a CONTEXT type index on the BLOB column:

```
CREATE INDEX plpdf_rinst_data_idx ON plpdf_rinst(rep_data) INDEXTYPE IS CTXSYS.CONTEXT;
```

Performing Queries

To prove your indexes were created properly and to demonstrate a number of ways for querying your content management repository, run through a number of test cases:

The following simple example demonstrates returning documents that match one specific word. In this case, you are searching for the word "oracle" in any of your documents:

```
SELECT SCORE(1) SCORE, ID, REP_ID, RUN_END  
FROM plpdf_rinst_data_idx  
where run_status = 'F'and contains(rep_data,'oracle');
```

```
SCORE ID REP_ID RUN_END  
---- -- -  
100 345 17 12-FEB-05 25 678 22 13-FEB-05
```

CTXCAT Indexes



A CTXCAT index is an index optimized for mixed queries. You can create this type of index when you store small documents or text fragments with associated structured information. To query this index, you use the CATSEARCH operator and specify a structured clause, if any. Query performance with a CTXCAT index is usually better for structured queries than with a CONTEXT index. To achieve better performance, your CTXCAT index must be configured correctly.

Query Applications

The purpose of a query application is to enable users to find documents that contain one or more of the search terms. For example, an application with a browser interface might enable users to query a company Web site consisting of HTML files, returning those files that match a query. After executing a query, query applications typically present a hit list of all documents that satisfy the query along with a relevance score. This list can be a list of PL/PDF reports titles.

Your application presents a hitlist in one or more of the following ways:

- By showing documents ordered by score
- By showing structured fields related to a document
- By showing document hit count