

# Indexing Arabic Documents with Oracle Universal Content Management and Oracle Text 10g/11g

This brief document describes how Oracle Content Server configured with Oracle Inbound Refinery and Oracle Text 10g/11g can be set up to index Arabic content. The information contained in this document is only relevant to versions of Content Server 10gR3 before 10.1.3.3.3. Content Server 10.1.3.3.3 introduces a new configuration parameter to allow native document to be indexed.

## Overview

Oracle Content Server 10g configured with Oracle Text 10g/11g might not index converted Arabic PDFs properly. Additional configuration, including modification of an `Idoc` resource include, is required for proper indexing of Arabic documents.

## How to Use this Capability

In order to set up Content Server 10g to properly index Arabic PDF documents using Oracle Text 10g/11g, the following steps need to be taken:

1. Set `IsAllowIndexVault` configuration variable to true in the `config.cfg` file. This will cause the Content Server to include vault file information for later evaluation by `Idoc` resource includes.
2. Modify the `std_indexable_format_script` resource include and pick out vault files to be converted/indexed appropriately. The proper method to do this is to write a component that overrides the resource include. A sample component, `UseNativeFileInIndex`, that shows how to do this is available at <http://www.oracle.com/technology/products/content-management/ucm/samples/index.html>. This sample component demonstrates how to pick out vault files as the preferred indexing rendition.

## Configuration in Content Server 10.1.3.3.3

In Content Server 10.1.3.3.3, there is a new configuration variable, `UseNativeFormatInIndex`. Setting this configuration value to true, will cause the content server to index the vault file by default. While it is similar to the `IndexVaultFile` configuration capability, unlike that configuration setting, `UseNativeFormatInIndex` does not prevent the creation of normal weblayout renditions. There is no need to modify the `std_indexable_format_script` resource include.