# Endeca Content Acquisition System

## Documentum Content Server Connector Guide

**Version 3.0.2 • March 2012**

ENDECA

# Contents

**Preface.....7**
About this guide.....7
Who should use this guide.....7
Conventions used in this guide.....8
Contacting Oracle Endeca Customer Support.....8

**Chapter 1: Configuring the CAS Server for Documentum.....9**
Documentum Content Server versions supported by this connector.....9
Setting up the CAS Server for Documentum Content Server.....9
Configuration properties for the Documentum Content Server connector.....10
Permission mapping.....11

**Chapter 2: Limitations of the Documentum Content Server Connector.13**
Limitations.....13

# Copyright and disclaimer

Copyright © 2003, 2012, Oracle and/or its affiliates. All rights reserved.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners. UNIX is a registered trademark of The Open Group.

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

This software or hardware and documentation may provide access to or information on content, products and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

Rosette® Linguistics Platform Copyright © 2000-2011 Basis Technology Corp. All rights reserved.

Teragram Language Identification Software Copyright © 1997-2005 Teragram Corporation. All rights reserved.

# Preface

Oracle Endeca's Web commerce solution enables your company to deliver a personalized, consistent customer buying experience across all channels — online, in-store, mobile, or social. Whenever and wherever customers engage with your business, the Oracle Endeca Web commerce solution delivers, analyzes, and targets just the right content to just the right customer to encourage clicks and drive business results.

Oracle Endeca Guided Search is the most effective way for your customers to dynamically explore your storefront and find relevant and desired items quickly. An industry-leading faceted search and Guided Navigation solution, Oracle Endeca Guided Search enables businesses to help guide and influence customers in each step of their search experience. At the core of Oracle Endeca Guided Search is the MDEX Engine,™ a hybrid search-analytical database specifically designed for high-performance exploration and discovery. The Endeca Content Acquisition System provides a set of extensible mechanisms to bring both structured data and unstructured content into the MDEX Engine from a variety of source systems. Endeca Assembler dynamically assembles content from any resource and seamlessly combines it with results from the MDEX Engine.

Oracle Endeca Experience Manager is a single, flexible solution that enables you to create, deliver, and manage content-rich, cross-channel customer experiences. It also enables non-technical business users to deliver targeted, user-centric online experiences in a scalable way — creating always-relevant customer interactions that increase conversion rates and accelerate cross-channel sales. Non-technical users can control how, where, when, and what type of content is presented in response to any search, category selection, or facet refinement.

These components — along with additional modules for SEO, Social, and Mobile channel support — make up the core of Oracle Endeca Experience Manager, a customer experience management platform focused on delivering the most relevant, targeted, and optimized experience for every customer, at every step, across all customer touch points.

## About this guide

This guide describes the tasks necessary to configure the Documentum Content Server CMS connector.

It assumes familiarity with the concepts of the Endeca Content Acquisition System and the Endeca Information Transformation Layer. For more information, see the *Endeca CAS Developer's Guide* and the *Endeca Forge Guide*.

## Who should use this guide

This guide is intended for application developers who are building applications using the Endeca Content Acquisition System, and are responsible for gathering, crawling, joining and feeding the data in different source formats into the Endeca pipeline to transform them into Endeca records.

# Conventions used in this guide

This guide uses the following typographical conventions:

Code examples, inline references to code elements, file names, and user input are set in `monospace` font. In the case of long lines of code, or when inline monospace text occurs at the end of a line, the following symbol is used to show that the content continues on to the next line: ¬

When copying and pasting such examples, ensure that any occurrences of the symbol and the corresponding line break are deleted and any remaining space is closed up.

# Contacting Oracle Endeca Customer Support

Oracle Endeca Customer Support provides registered users with important information regarding Oracle Endeca software, implementation questions, product and solution help, as well as overall news and updates.

You can contact Oracle Endeca Customer Support through Oracle's Support portal, My Oracle Support at *https://support.oracle.com*.

Chapter 1

# Configuring the CAS Server for Documentum

Set up the CAS Server for Documentum Content Server, and set Content Server-specific options in the CAS Console. See the "Endeca CAS API Guide" for details on crawling a Documentum Content Server repository through the CAS API.

## Documentum Content Server versions supported by this connector

The Documentum Content Server connector supports the following versions of Documentum Content Server repositories:

- Documentum Content Server 5.3
- Documentum Content Server 6.0
- Documentum Content Server 6.5

**Note:** Documentum Content Server 5.2.5 and earlier is not supported.

## Setting up the CAS Server for Documentum Content Server

To crawl a Documentum Content Server repository (docbases), perform additional configuration steps for the CAS Server.

In order to crawl a Content Server repository, the CAS Server needs to have access to the Documentum shared libraries.

**Note:** The following procedure uses the Documentum Foundation Classes 6.x. These classes are compatible with Documentum 5.3 and later.

**To set up a CAS Server for Documentum:**

1. Install and configure the Documentum Foundation Classes 6.x libraries using the Documentum installer. Refer to the Documentum documentation.

2. Copy the following file from the Documentum Foundation Classes installation to the `CAS\version\lib\cas-server-plugins\entropysoft` directory (on Windows) or `CAS/version/lib/cas-server-plugins/entropysoft` directory (on UNIX):
   - `dfc.jar` including any JAR files that `dfc.jar` depends upon. (You can identify dependencies by opening the manifest file for `dfc.jar` and copying the files listed in the manifest.)

   **Note:** By default, these DFC files are located under `C:\Program Files\Documentum\Shared` (on Windows).

3. Copy `dfc.properties` from the Documentum Foundation Classes installation directory to `CAS\workspace\conf` (on Windows).

   **Note:** By default, this file is located under `C:\Program Files\Documentum\config` (on Windows).

4. Restart the Endeca CAS Service.

Now the CAS Server is set up to communicate with the Documentum repository.

In order to crawl a Documentum repository, you also set data source options in the CAS Console.

# Configuration properties for the Documentum Content Server connector

To configure a Documentum Content Server connector, specify the configuration properties listed below.

**Note:** In addition to configuring the connector-specific properties listed below, you must enter values for the data source username and password.

Create the following configuration properties using either CAS Console or the CAS Server Command-line Utility.

| CAS Property Display Name | CAS Property Name | Property Description |
|---|---|---|
| **Docbase Name** | `docbase` | (Required). Specify the name of the Documentum repository. |
| **Webtop URL** | `webtopUrl` | (Optional). The base URL of the local Documentum StringWebtop installation. For example: `http://myhost:8080/webtop/`. |
| **User Name** | `username` | (Required). Specify a valid username that has access to the Documentum repository. |
| **Password** | `password` | (Required). Specify a password for the above username. |

| CAS Property Display Name | CAS Property Name | Property Description |
|---|---|---|
| **Domain** | `domain` | (Optional). Specify the domain of the user name. |

**Note:** Properties are case sensitive.

# Permission mapping

The following table shows the mapping between Documentum permissions and the resulting Endeca record properties are produced.

| Documentum permission | Endeca record properties |
|---|---|
| Browse | `Endeca.CMS.AllowReadProperties` |
| Read | `Endeca.CMS.AllowReadContent`<br>`Endeca.CMS.AllowReadProperties` |
| Relate | `Endeca.CMS.AllowReadContent`<br>`Endeca.CMS.AllowReadProperties` |
| Version | `Endeca.CMS.AllowReadContent`<br>`Endeca.CMS.AllowReadProperties` |
| Write | `Endeca.CMS.AllowReadContent`<br>`Endeca.CMS.AllowReadProperties` |
| Delete | `Endeca.CMS.AllowReadContent`<br>`Endeca.CMS.AllowReadProperties` |

Chapter 2

# Limitations of the Documentum Content Server Connector

This section describes limitations of the Documentum Content Server connector.

## Limitations

### Renditions

Renditions are currently not supported. Only the original content can be retrieved or updated by the connector. The "renditions" document property contains the list of available rendition formats for a given document.