

ORACLE OUTSIDE IN TECHNOLOGY

KEY FEATURES AND COMPONENTS

LEVERAGE ORACLE OUTSIDE IN TECHNOLOGY TO:

- Extract and scrub hidden data from Microsoft Office and PDF files
- Accurately identify over 500 file types
- Extract text and metadata without 3rd party applications
- Convert files into different output formats including PDF, XML, HTML or TIFF
- Provide a high-fidelity view of file contents

ORACLE OUTSIDE IN TECHNOLOGY SDKs

- Oracle Outside In Clean Content
- Oracle Outside In File ID
- Oracle Outside In Content Access
- Oracle Outside In Search Export
- Oracle Outside In HTML Export
- Oracle Outside In XML Export
- Oracle Outside In PDF Export
- Oracle Outside In Image Export
- Oracle Outside In Viewer

Oracle Outside In Technology enhances hundreds of software applications from the world's leading technology companies for search, e-mail, e-discovery, text mining, content management, legal, security, compliance, archiving, data forensics, collaboration, and many others. Oracle Outside In Technology provides software developers with a comprehensive solution to access transform and control the contents of nearly 500 unstructured file formats. From standard business documents to specialty formats and legacy files, Oracle Outside In Technology turns unstructured files into accessible information.

Oracle Outside In Technology Software Development Kits

Oracle Outside In Technology is organized into nine software development kits (SDKs). Each SDK is optimized to provide a particular file access, transformation or control solution, but they are all highly flexible and interoperable. This interoperability allows powerful solutions to be built by combining the individual SDKs. For example, text extracted using one of the search-enabling SDKs can be used to generate annotations or redactions in the output of an export SDK.

Developers can quickly implement any combination of Oracle Outside In Technology SDKs to provide exactly the right functionality in their application while minimizing integration effort and code footprint. The SDKs offer a wide range of options to give developers programmatic control of their workflow and output. Thorough documentation and sample applications with source code are included to further accelerate implementation.

Oracle Outside In Clean Content

Outside In Clean Content addresses particularly challenging issues in native file processing. Focusing specifically on widely used formats (Microsoft Office and PDF), it extracts all text, properties, hidden information and system data embedded in native files. Clean Content can also programmatically modify native files enabling features such as scrubbing, property modification and document assembly. For example, a recent implementation leveraged Clean Content to “burst” presentation slide decks into individual slides, from which users can select to create an entirely new slide deck. Outside In Clean Content is a pure Java technology that offers Java, C/C++ and .NET APIs.

MARKET LEADERSHIP

ORACLE OUTSIDE IN TECHNOLOGY IS THE CLEAR MARKET LEADER FOR FILE EXTRACTION, CONVERSION AND VIEWING. IT IS USED BY:

- 8 of 10 leading Electronic Discovery software providers
- 8 of 10 leading Electronic Discovery service providers
- 10 of 15 leading Information Access providers
- 5 of 5 ECM leaders
- 7 of 7 Portal leaders
- And hundreds of others

Oracle Outside In File ID

Oracle Outside In File ID identifies over 500 file types without using unreliable file extensions or MIME types. The technology is particularly useful at the start of a workflow when dealing with unknown data, in security applications, and when the format of a file needs to be quickly and accurately identified. Outside In File ID now supports the identification of several multimedia formats, allowing those file types to be routed appropriately through workflow or flagged for special handling.

Oracle Outside In Content Access

Oracle Outside In Content Access extracts text and metadata from nearly 500 file types and automatically translates it from multiple possible encodings into a single encoding specified by the developer. It is optimized for performance, interactively providing data to the host application in memory as the input file is processed. The technology is widely used in search, data forensics, and security applications.

Oracle Outside In Search Export

Oracle Outside In Search Export provides the text, metadata, and pagination information of the supported file types in an XML, HTML, or text specifically designed for search and forensic applications. The technology differs from Oracle Outside In Content Access in that the output is generated as a stream or file, rather than exchanged interactively. It supports nearly 500 file types and offers the developer a number of options, including a choice of four output formats:

- **SearchML**—Lightweight XML containing text, embeddings, and metadata optimized for search and text extraction
- **SearchHTML**—HTML optimized for Web crawlers but with limited display formatting
- **SearchText**—Plain text file (UTF-8 encoded Unicode) with properties and body text from the input file
- **PageML**—XML that provides paginated text

Oracle Outside In HTML Export

Oracle Outside In HTML Export converts the contents of nearly 500 file types into HTML, rendering embedded graphics as a choice of GIF, JPEG, or PNG. Oracle Outside In HTML Export uses sophisticated template architecture, as well as a rich set of options, to allow a high degree of customization of the HTML output. When used in combination with Oracle Outside In Content Access or Oracle Outside In Search Export, annotations or redactions can be added to the HTML output programmatically. Oracle Outside In HTML Export can use cascading style sheet formatting for more-complex formatting, or alternatively can be scaled down to use least-common-denominator versions of HTML for the broadest compatibility with browsers on various platforms or devices. Its ability to provide browser access to hundreds of file types without plug-ins or other proprietary applications makes it one of the most widely used SDKs of the Oracle Outside In Technology suite.

VALUE PROPOSITION

GAIN COMPETITIVE
ADVANTAGE

- Extend your application's reach and value with the ability to access, transform and control unstructured information

REDUCE TIME TO MARKET

- Quickly and easily implement new features with robust and flexible development tools

DO MORE FOR LESS

- Reduce engineering time and expense by leveraging market-leading technology at a fraction of its development and maintenance cost

Oracle Outside In XML Export

Oracle Outside In XML Export converts and normalizes the content of supported file types into XML that closely mirrors the information in business documents. In addition to the document content itself, FlexionDoc includes information about document properties, structure, and formatting. The resultant XML can be consumed by an application or further transformed by the developer via XSLT into a developer-defined schema. Oracle Outside In XML Export is appropriate for use in any application that can benefit from normalizing documents into XML, which explicitly describes all the elements of the document's content, structure, properties, and formatting.

Oracle Outside In PDF Export

Oracle Outside In PDF Export is the first cross-platform, application-independent PDF conversion solution for the OEM software market. It includes its own rendering engine so it does not require native applications to generate PDF, which simplifies workflow and decreases cost. It is interoperable with other Outside In Technology SDKs such that the offset information of text extracted using one of the search enabling SDKs can be used to programmatically annotate text in the PDF output. PDF Export can generate fully compliant PDF/A-1a (Archive) PDF, which is a significant added value for applications in the archive and compliance markets.

Oracle Outside In Image Export

Oracle Outside In Image Export converts the supported file types into TIFF, JPEG, JPEG 2000, BMP, GIF, or PNG images. Oracle Outside In Image Export offers numerous options, including the ability to size the image output from thumbnail to full size and control image resolution. When used in combination with Oracle Outside In Content Access or Oracle Outside In Search Export, annotations can be added to the output programmatically. Oracle Outside In Image Export is used extensively in e-discovery and in other applications that require high fidelity, static images of business documents.

Oracle Outside In Viewer

Oracle Outside In Viewer renders high-fidelity representations of files without using the files' native applications. It enables client-side file viewing, printing, and copy/paste functionality of nearly 500 file formats. An annotation API is also included, which allows the developer to highlight, hide, or insert text in a document "view" without altering the underlying native file. It is also available as an ActiveX control for Microsoft Visual Basic or other ActiveX-compliant development environments.

Contact Us

For more information about Oracle Outside In Technology], please visit oracle.com or call +1.800.ORACLE1 to speak to an Oracle representative.



Copyright © 2009, Oracle and/or its affiliates. All rights reserved.

This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners. 0109