

# ORACLE OUTSIDE IN TECHNOLOGY

## KEY FEATURES & COMPONENTS

### LEVERAGE ORACLE OUTSIDE IN TECHNOLOGY TO

- Extract and scrub hidden data from Microsoft Office and PDF files
- Accurately identify over 400 file types
- Extract text and metadata without 3<sup>rd</sup> 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
- Oracle Outside In PDF Export
- Oracle Outside In Image Export
- Oracle Outside In Viewer Technology

*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 over 400 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 offers a different file access, transformation or control solution and is optimized to solve a particular problem; but all are highly flexible and interoperable, generally sharing the same code base and naming convention. This interoperability allows powerful solutions to be built by combining Oracle Outside In technologies. 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. Its extended extraction includes the ability to analyze and process malformed documents, which is critical to accurate text extraction from PDFs. Clean Content can also programmatically modify native files enabling features such as scrubbing, property modification and document assembly. Outside In Clean Content is a pure Java technology that offers Java, C++ and .NET APIs.

### **Oracle Outside In File ID**

Oracle Outside In File ID identifies file types without using unreliable file extensions or MIME types. It has long been a critical component of Oracle Outside In Technology but is now available as a separate SDK. 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 identified.

### **Oracle Outside In Content Access**

Oracle Outside In Content Access extracts text and metadata from more than 400 file types and automatically translates the text and properties 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 and data forensic applications and is designed for high-throughput server environments.

### **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 form 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 more than 400 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 more than 400 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, SearchML, 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.

**Oracle Outside In XML Export**

Oracle Outside In XML Export converts and normalizes the content of supported file types into XML defined by Oracle's FlexionDoc schema. This schema uses XML tags designed to closely mirror the information in files created in popular business applications. In addition to the document content itself, FlexionDoc includes information related to document properties, structure, and formatting. The resultant XML can be consumed by an application or further transformed by the developer via Extensible Style Language Transformations 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 apply hit-highlighting in the PDF output. It supports the most popular PDF conversion features including: linearized PDF for fast browser display; automatic generation of XMP metadata; and, numerous user options such as page range and size specification, programmatic addition of non-transparent watermarks, output resolution control, and many others.

**Oracle Outside In Image Export**

Oracle Outside In Image Export converts the supported file types into TIFF, JPEG, BMP, GIF, or PNG images. As with all Oracle Outside In Technology SDKs, 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, redactions, or SearchML can be added to the output programmatically. Oracle Outside In Image Export is used in e-discovery and in other applications that require access to high-fidelity, static images of business documents.

**Oracle Outside In Viewer Technology**

Oracle Outside In Viewer Technology renders high-fidelity representations of files without using the files' native applications. It enables client-side file viewing, printing, and copy/paste functionality of more than 400 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. Its use is appropriate for solutions that can benefit from desktop viewing functionality. It is also available as an ActiveX control for Microsoft Visual Basic or other ActiveX-compliant development environments.

**ORACLE OUTSIDE IN  
TECHNOLOGY: KEY  
BENEFITS**
**A BUILD VS. BUY  
DECISION**

Hundreds of technology companies have recognized that when it comes to file transformation, it's more sensible to buy rather than to build. With more than a decade of research and development and substantial ongoing investment to support emerging architectures, file formats, and operating systems, Oracle Outside In Technology provides significant value at a fraction of its real development and maintenance cost.

**Companies Using Oracle Outside In Technology**

AdLib eDocument Solutions	Fios	Postini
BrightPlanet	Global 360	Recommind
BEA	Guidance Software	Reconnex
Canon	IBM	Resume Mirror
Captaris	Interwoven	Ricoh
CaseCentral	InQuira	RightNow Technologies
CaseData	Inxight Software	SAIC
Cerulean	IronPort Systems	SER Storage Technology
Chiliad Publishing	KPMG Forensic	SiteScape
Cimage NovaSoft	Kroll Ontrack	Stratify
Code Green Networks	LexisNexis	Sun Microsystems
Convera	Lockheed Martin	Thales Communications
Digitech Systems	Lucid IQ	Thomson West
Discovery Mining	McAfee	WebMD
DocuLex	Merrill Corp.	Westbrook Technologies
Electronics for Imaging	Motorola	World Software
Endeca Technologies	Northrop Grumman	X1 Technologies
Engenium	Novell	Xerox
Epiq Systems	Open Text	Xiotech
Fast Search & Transfer	Hummingbird	Yahoo!
FileNet	Orchestria	
	PerfectLaw Software	

**Contact Us**

For more information on Oracle Outside In Technology, call +1.800.ORACLE1 to speak to an Oracle representative or visit [oracle.com/products/middleware](http://oracle.com/products/middleware).

Copyright © 2008, 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 is it 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.