

Oracle9iAS Portal – Web Clipping Portlet

An Oracle Technical White Paper
October 2002

Oracle9iAS Portal – Web Clipping Portlet

Introduction	3
Oracle9iAS Portal -- Web Clipping Portlet	3
Supported Page Content	6
Enhanced Performance with Caching	7
Secure Communication with Portal.....	7
Conclusion.....	8

Oracle9iAS Portal – Web Clipping Portlet

INTRODUCTION

The benefits of today's enterprise portals are well known – by consolidating information from various corporate Web sites and access to back-end applications, tailored for a particular user, intranet content is better organized and searched, traditional, paper-based processes can be eliminated, and users save both time and effort in getting to needed resources. But what about the job of the portal page designer? He/she may need to sift through hundreds or thousands of Web sites to find appropriate content, and can spend hundreds or thousands of man hours doing so for a medium to large-sized business. The task can be extremely laborious, and much time, money, and effort can be spent to custom build a portal, just to find out that when content needs to be replaced or added, or application interfaces change, the work needed could be just as significant as the original undertaking. What is needed are intuitive, straightforward tools to speed the creation of portlets using existing Web-based content and functionality.

ORACLE9iAS PORTAL -- WEB CLIPPING PORTLET

The Web Clipping Portlet, a new feature in the November Portal Developer's Kit (PDK), offers an easy, intuitive way to capture content and functionality from existing Web sites and present them as portlets.

Web Clippings are pieces of existing Web content that can be re-purposed in other Web pages, particularly portals. The simplest form of Web Clippings can be sections of HTML code in a Web page; clippings can also include non-HTML elements within a Web page, such as applets or JavaScript code.

The Web Clipping architecture, relative to Portal, is shown in Figure 1.

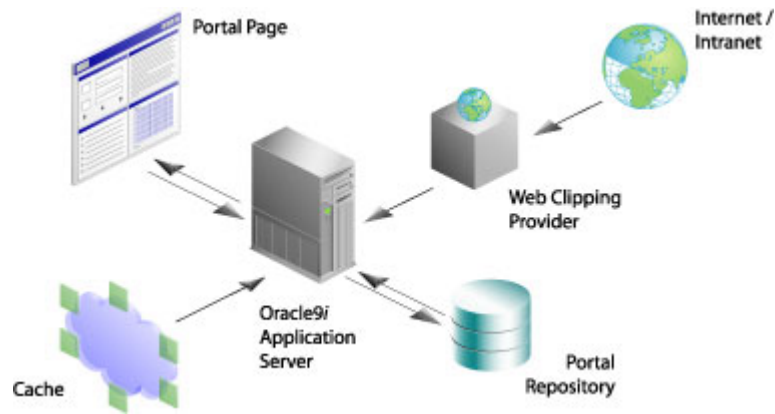


Figure 1 Web Clipping Architecture

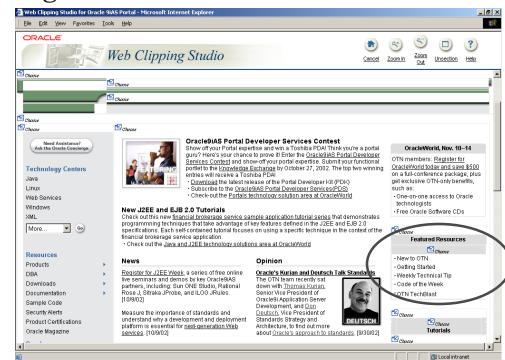
The Web Clipping Provider is a fully compliant Portal Web Provider, “providing” all the information for its portlets, fetching external content, and driving portlet operations. The provider’s schemas can be installed in an existing Oracle9i (9.0.1) or later database; in this paper, we will refer to this database as the *Web Clipping Repository*. In addition, the provider drives the creation of Web clippings through the Web Clipping Studio tool; the tool starts when the Portal user customizes a Web Clipping Portlet or edits the default portlet settings.

The studio allows the user to navigate to a Web page, similar to a standard browser, divide the page into “clipped” sections, and choose the appropriate section for presentation in the portal. Behind the scenes, the studio records all the URLs directly involved in getting to the source Web page, and “replays” the navigation upon portlet “show” mode.

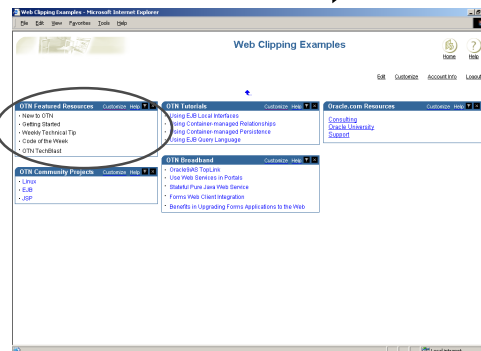
For example, the basic flow using the Web Clipping Studio to capture the ‘Featured Resources’ section of the Oracle Technology Network (OTN) homepage, is shown in Figure 2.



1. Browse to Source Web Page



2. 'Section' Page, Choose Clipping



3. Portal Page with Web Clipping

Figure 2 Using the Web Clipping Studio to Capture OTN Content

A common issue with traditional “screenscraping” technology is coping with changes in style and position of the scraped content within the source. Once the

source changes, the screenscraping script must also change to retrieve the right content. The Web Clipping Portlet addresses this issue by storing characteristic features of each clipping. For example, if a clipped table moves from one place to another in the source page, the Web Clipping engine can find the table again using the internal “fuzzy match” algorithm; or, say the source for a piece of clipped text changes in font, size, or style – the engine is resilient enough to locate the right clipping, regardless of these changes.

Because the Web Clipping Provider is a Portal Web Provider, it can be hosted on a remote Web server and managed outside Portal; the provider machine can also be scaled, accordingly, to meet the demand of Web Clipping Portlet requests. Refer to the PDK documentation (<http://portalstudio.oracle.com>) for more information on Web Providers.

SUPPORTED PAGE CONTENT

The Web Clipping Studio supports clipping of HTML 4.0.1 pages, including:

- Clipping of <table>, <td>, , , <div> tagged content
- Preservation of <head> styles and fonts, Cascading Style Sheets (CSS)
- UTF-8 compliant character sets
- Navigation through hyperlinks (HTTP GET), form submissions (HTTP POST), frames, URL redirection.

Form submissions are securely recorded for “replay” on portlet “show” mode. HTTPS-based external Web sites can also be navigated and clipped, provided that appropriate server certificates are acquired. Refer to the *Oracle9iAS Portal Web Clipping Portlet User's and Administrator's Guide* for more information.

- Portal user customization of input parameters
- Session maintenance for Web sites that require login.

Specifically, the Web Clipping Provider maintains session to external Web sites, and to the Portal instance on portlet “show” mode.

In addition to form-based submission of username/password information, HTTP “Basic” and “Digest” Authentication Schemes are supported. For example, the studio allows navigation through HTTP authentication challenges that pop up in browser dialog boxes.

Most non-HTML elements are also supported, including applets, plugin content (e.g. embedded Quicktime video, Macromedia Flash presentation), and client-side Javascript (v1.2).

ENHANCED PERFORMANCE WITH CACHING

To accelerate portlet and associated metadata retrieval, the Web Clipping Provider can use either Portal's validation-based caching, or configured to use Oracle9iAS Web Cache.

By default, the Web Clipping Provider utilizes Portal's validation-based caching for all of its Web Clipping Portlets. With this type of caching, the Parallel Page Engine (PPE) always contacts the Web Clipping Provider to determine if the cached item is still valid. If so, the PPE retrieves content from the disk-based Portal Cache; if the cached item is not valid, the Web Clipping Provider must initiate a request to retrieve content from the external Web site.

Alternatively, the Web Clipping Provider can be configured to take advantage of the Oracle9iAS Web Cache; this invalidation-based cache can be used to store portlet content, in-memory, between the Portal instance and the provider. With this type of caching, an item remains in the cache until the cache receives notification that the item needs to be refreshed. For example, if the Web Clipping Portlet contains content that is updated on a regular basis, the cached content will also be invalidated on a regular basis. Compared to the default Portal validation-based caching, invalidation-based caching, as shown in Figure 3, decreases the number of requests the Web Clipping Provider must entertain while maintaining the same network traffic per round trip involving PPE. Depending on your deployment scenario, you may prefer using one caching method over the other. For more information about caching, see *Oracle9iAS Portal Configuration Guide for Release 2 (9.0.2)*.

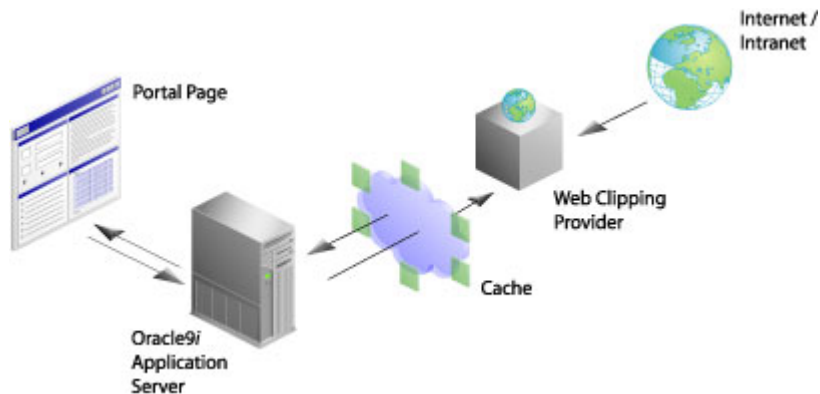


Figure 3 Invalidation-based Caching Provided by Web Cache

SECURE COMMUNICATION WITH PORTAL

In addition to standard Oracle database encryption of clipping and related metadata in the Web Clipping Repository (using DBMS_OBFUSCATION_TOOLKIT), the Web Clipping Provider can also be configured to use Oracle Advanced Security Option (ASO), if licensed, with the

Web Clipping Repository. This provides DES-based encryption of communication between the provider and Web Clipping Repository. See the *Oracle Advanced Security Administrator's Guide* for more information about this option.

CONCLUSION

Oracle9iAS Portal stands at the forefront of portal technology, offering an open, productive, and complete framework on which to build robust, rich, and personalized portals.

With the Web Clipping Portlet, Portal page developers can easily and quickly extract Web-based content for portlet presentation. The Web Clipping Portlet allows Portal Page Designers to create Web clippings on the fly by simply navigating to the source Web page in a browser-like fashion, and selecting the portion of the page to clip. Web clippings greatly simplify aggregation of Web content into a portal. For example, using Web clippings, large organizations with hundreds of Web sites scattered throughout their enterprise can easily consolidate content into a centralized portal.

The Web Clipping Portlet allows the reuse of a variety of Web content, including the basic support of pages written with HTML 4.0.1, JavaScript, applets, and plug-in enabled content, as well as cookie-based session management, through HTTP GET and POST (form submission). The portlet supports navigation through various styles of login mechanisms, including form and JavaScript-based submission, through HTTP Basic and Digest Authentication. In addition, "fuzzy matching" of clippings is supported; this means if a clipping gets reordered within the source page or if its character font, size, or style changes, it will still be identified correctly by Portal and delivered as the Portlet content.



Oracle9iAS Portal – Web Clipping Portlets
October 2002

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
www.oracle.com

Oracle Corporation provides the software
that powers the internet.

Oracle is a registered trademark of Oracle Corporation. Various
product and service names referenced herein may be trademarks
of Oracle Corporation. All other product and service names
mentioned may be trademarks of their respective owners.

Copyright © 2000 Oracle Corporation
All rights reserved.