WSC^{*} > Semantic Web Use Cases and Case Studies

Case Study: Enhancing Content Search Using the Semantic Web

Mike DiLascio, Siderean Software, and Justin Kestelyn, Oracle Corporation

March 2007



General Description

The Oracle Technology Network (OTN) is the main source of technical information for the Oracle developer community. The Web site provides access to blogs, podcasts, discussion fora, product documentation, notifications of product releases, and software downloads. The dynamism, richness, and complexity of this information has made it difficult for traditional search techniques to provide access to data of interest.

Siderean Software and Oracle have worked to apply Semantic Web technologies to make data discovery and navigation more effective. The solution is based on the integration of Oracle Secure Enterprise Search and Siderean's Seamark Navigator. The Web site is available at: http://otnsemanticweb.oracle.com

This approach improves information access by aggregating many sources of content through a single portal. This approach enables a unified interface for searching and browsing the rich multi-media data, which is valuable in helping users to identify all information of interest with a single query. Users can also personalize the environment by selecting feed items of interest and by specifying a preferred layout for the information. To help users find the precise information of interest, the system provides visualization approaches such as tag clouds, contributor clouds, and timelines. Figures 1 and 2 show screen snapshots of the enhanced OTN Web site.

To keep the Web site current, the application pulls content of interest from multiple RSS feeds every hour. Seamark enhances the metadata provided by the RSS feed by using a combination of pattern-matching techniques to identify the subject matter of an item. Once the subject has been established, terms from an Oracle proprietary taxonomy are used to record the concepts and entities to which the item refers. As terms for the concepts and entities in metadata are standardized, feeds constructed using this technology tend to be more targeted, with fewer inappropriately included items. Since the same approach of associating metadata is used for all media types, these feeds can deliver relevant content from diverse media platforms.

This Semantic Web-based approach provides the annotation required to support a rich discovery experience based on dynamic navigation of entities and their relationships. Since terms are taken from an ontology, it becomes possible to group content that relates to the same parent concepts. This process, in turn, makes it easier for a user to comprehend many results in search. It also makes it easier to narrow results to a precise topic of interest.

Integration and deployment of the solution was very quick and easy as both Secure Enterprise Search and Seamark are available as Web Services. The OTN solution was deployed in approximately eight weeks. Future enhancements are planned to support additional data types, enable greater personalization, and extend visualization approaches.

	emantic Web + Hestila Firefan			14 IP 1
But you when Booken	ter Too Re			
			TIP CO	
			See on these last	nter (4
Citilitate Headhers B	nanikali funimi Arkani Pokashi Moge	Meet .	ture .	
anutif er	Welcome to OTN Semantic Web (Bet	a)		
New York Contraction (1991) Institute Contractions (1991) Marinese Contractions (1991) Marinese (1991) Marinese (1991) Marinese (1991)	The stand of the second state of the second st	641		
technik monory (sessible (12)	an and a result	annus	A Sec of Fands of an	class and view
euriteing di	and and a second	antits	and the states	
6 (f)	noted Chaine Resi, 13 and Dowin Romanne Lone works When the residence of the second state of the second st	2000年50.2995年-	State Incomentation and the Low off	
International States (III)		The Table of Farm with County Find	fernere 10, 100, 12, 13	
and be stope Task (MA)	Orderin SCA Style Beat/Franklass-Technolus/Autors profiled 2027-01-03	and the second se	provide of 2000-13-21 Crisis (Cylinates Unit Tig Remote 2.010.2.0.00 for Landon 3	
neria Pranson Malakenseria (*123) (m. Tadas (1947)	Lindian and Diskons.	Rule rour Dee Deale Archite freeson 2 Chatter in Unit Ref. 2006 (1):01 poster 2006 (1):01		
Place Are en	Formant Meet the Charle Linux nerver Teach product 2006-73-31	Interna dissiona Interiment		
reacture (100) and (37)	Divertiged Double (Sill, Devenues: 1.1 socied, 2006-13-15	Learning Crucks & C.F. Allegarter's Dist- posed 2008 (11:3)		
10 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C			production 2006 10 20	
Classic could be developed to (2021) developed to (2021)		Salate at the de	Canada (consection)	
Kalana (HTT)) karpropietaragonasti (1.01) milit Managamenti (111)	8.00	PORATE	MICHINA PARA	
en (21%) en (21%) en (21%)	Finish to hep "herbit" in D-Box #12 product 2007-01-24	Disk in Apple date Experies and its Unit Experience. The First per Party in Institute of 2007 01 42	forms Devi bits formation and moting edenal Securit Service marines	
KOMO (THE	Anex ROAmed besidepart, under all the ROA Decis		by definitive provided set (2007-01-08)	
er das Cherrikel Auslidentiver (213) Nord	Redge" andres (2007-01-04 Dear No. 1077, and before 1	Entrance Variant, April due Variagement Parts for Education Linds, Properties Comprise, and Bellet Parts of State of State	Onfor properties, information on out, 2020, Daystern, Diringation work (2022) in sourced 202201 protocol on 2020, 201 and	
418	Rest, Stat. Ser. and Landard.	Drace Reduces research Castal Distance Roos		
82 (FFFB) 88 (9988) 86 (227)	Amon Mile 527, articlastics." Acceler 2007-01-04	Investigation of States of States	Dani Jent Con In Druce and American In Higher and American possible in 1997 (1) and	
tie mit	Diskin mathematication and the second s		Entergence Lana dt., second Proce trabel	

Figure 1. This figure shows a screen snapshot of the Oracle OTN Web site. Portlets have been chosen by a user to support simultaneous navigation across many repositories. The panel on the left includes facets that are available for navigation, and indicate the number of items available for each link.

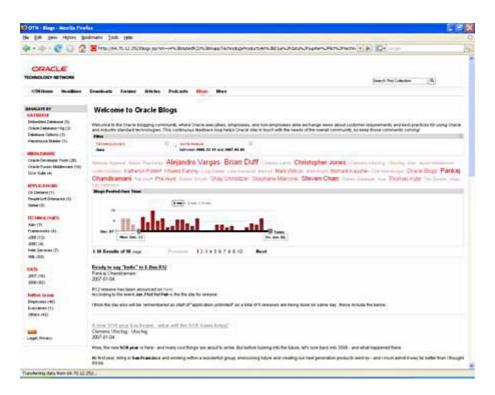


Figure 2. The Oracle OTN Web site supports many visualization capabilities along with Semantic Web search and navigation capabilities. This screen snapshot highlights the results of a search for blogs that include the term "Java", and were written between December 11, 2006, and January 5, 2007. The tag cloud indicates that Brian Duff and Alejandro Vargas have written a lot of these blogs. The bar chart shows the number of blogs that were written on each day over the

chosen time period. A list of the search results is highlighted near the bottom of the page. It is possible to view blogs according to dimensions such as time, author, and topic.

Key Benefits of Using Semantic Web Technology

- Enhanced search and navigation of a complex corporate Web site
- Improved visualization capabilities to guide users to data of interest
- Improved ability to personalize the Web site
- Enabled quick and simple implementation
- Stored data in RDF and enabled queries using SPARQL

© Copyright 2007, Siderean Software and Oracle Corporation