



Improve the way you create, manage and distribute information



**INNOVATION
INSPIRATION**

Creating an Enterprise Search Strategy

Integrating Navigation, Knowledge Management, and Traditional Search Technologies

Presented by Joe Peterson

www.innodata-isogen.com

The Challenge of Locating Information



- Organizations are facing an unprecedented level of information overload
- The amount of information available outstrips the ability of individuals to locate or track
- Information systems must adapt to assist users in finding the best information available

Knowing Where Information Exists



- At some level, someone tends to know what information is important
- Most electronic systems allow people to insert information, but rarely provides a consistent means to capture the “mental metadata” that exists in people’s mind
- Electronic systems instead tend to attempt to make guesses as to the nature of the information using a variety of techniques

Traditional Search Engines



- Search traditionally involves tools to scan files and create “electronic indices” keyed on search terms
- Powerful databases catalog all text and create rapidly accessible tables of information
- These tools are effective at some basic tasks but break down as the volume of information increases
- Proprietary search algorithms attempt to prioritize information for end users

Traditional Search Results

[Monitor Calibration](#)

Are things looking murky? Maybe you need to calibrate your monitor!
[epaperpress.com/monitorcal/](#) - 1k - [Cached](#) - [Similar pages](#)

[The USS Monitor Center | The Mariners' Museum](#)

Site deals with the discovery and recovery of the Ironclad, plus its conservation, history, museum, and online shop.
[www.monitorcenter.org/](#) - 21k - Oct 31, 2005 - [Cached](#) - [Similar pages](#)

[Christian Monitor: Home](#)

Christian Monitor provides resources and information on Christian persecution and the suffering Church. It encourages Christians to pray for persecuted and ...
[www.christianmonitor.org/](#) - 30k - Oct 31, 2005 - [Cached](#) - [Similar pages](#)

[WELCOME TO MONITOR. WE ARE TRUST FUND FREE.](#)

Monitor Records USA: Completely trust fund free. Made by blue collar Americans.
[www.monitorrecords.com/](#) - 3k - [Cached](#) - [Similar pages](#)

[Los Angeles International Airport - AirportMonitor - by Megadata ...](#)

January, February, March, April, May, June, July, August, September, October, November, December. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 ...
[www4.passur.com/lax.html](#) - 10k - [Cached](#) - [Similar pages](#)

[Monitor Outlet, Inc. - Your source for Plasma TV's, LCD Displays ...](#)

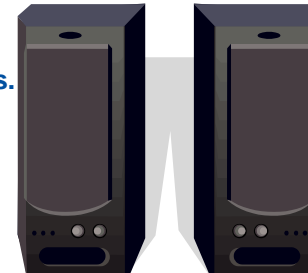
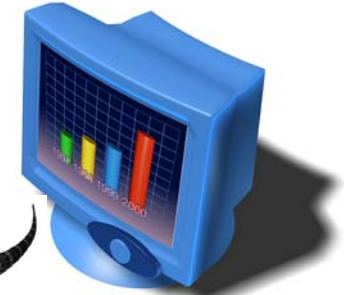
Monitors, LCD displays, plasma displays and projectors with complete online specifications and pictures.
[www.monitoroutlet.com/](#) - 35k - [Cached](#) - [Similar pages](#)

[DX Monitor](#)

Shareware to monitor OH2BUA Internet DX Cluster. ... DX Monitor is a standalone Windows program which monitors the DX announcements available on the ...
[www.benlo.com/dxmon.html](#) - 10k - [Cached](#) - [Similar pages](#)

[BBC Monitoring](#)

Includes news, information and comment gathered from the mass media around the world for service subscribers.
[www.monitor.bbc.co.uk/](#) - [Similar pages](#)



A Sea of Results

- Searching on an index may provide the results – but the results may be buried in a sea of information
 - In the previous example, a search for “dramatic male opera roles” yielded 603,000 results
- Proprietary algorithms to rank results may work for general use
 - Typically, they cannot indicate what is the “best” information
 - Multiple versions of the same documents may not be differentiated
 - Searching for “Innodata Isogen” yields 28,400 results, with the top being our corporate Web site

Algorithms are Limited in Applicability



- Search algorithms may provide some good results for certain uses
 - “Most popular” algorithms may help me find a corporate Web site
 - “Most popular” information may equate to mediocrity in a competitive business setting
- Keywords may assist, but are also limited in scope and applicability
 - Do all your users have the same vocabulary?
 - Keywords do not provide a useful means of browsing information

Traditional Search Traps



- Without knowing the details of the algorithm, you are subject to another corporation's control over what you find
 - A “find what is most popular” algorithm may conflict with my goals (i.e. finding what will distinguish me from others)
- Search engines may provide some configuration options, but typically your investment in these options are limited to a single vendor's software
 - Metadata can be added to assist the engine in organizing information
 - Information can be placed into taxonomies to allow some categorization, for those familiar with the taxonomy
 - Investment in these models means tying your knowledge to a single vendor's vision

Automatic Search is Insufficient



- Search engines cannot always identify the “best” information
 - Often multiple copies of similar information exists without indication of what is current or best
 - Old information may appear as “more relevant” than new information
- Experts close to the information often can identify what is “best” for a particular use

Knowledge is as Critical as Information



- Gathering information into standard forms is now recognized as a goal in most organizations
 - XML and other open formats exist to facilitate this move
- Knowledge is as critical to success as information
 - New open standards have emerged to address this need: i.e. XML Topic Maps, RDF
- Information technology is now attempting to leverage the experience gained in the library sciences
 - For too long our systems have relied on “brute force” processing power and ignored thousands of years of library research

Knowledge Hides in our Computer Systems

- Traditionally, computer applications scattered our knowledge of how information is used in arcane, proprietary single-use structures
 - Web site navigation bars
 - Menu structures in computer applications
- This knowledge is helpful to a single application, but cannot typically be used by other applications
 - Search engines generally cannot make sense of complex relationships by examining hyperlinks, menu bars or navigation structures
- Where two applications “meet” (i.e. relate to the same topic) there typically is not a way to definitively merge different frameworks

Capturing Knowledge

- Capturing knowledge requires computer science to look at the philosophical study of knowledge
- The term “ontology” is used to visualize this study

“The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D .”

John Sowa, <http://www.jfsowa.com/ontology/>

Electronic Standards for Ontologies



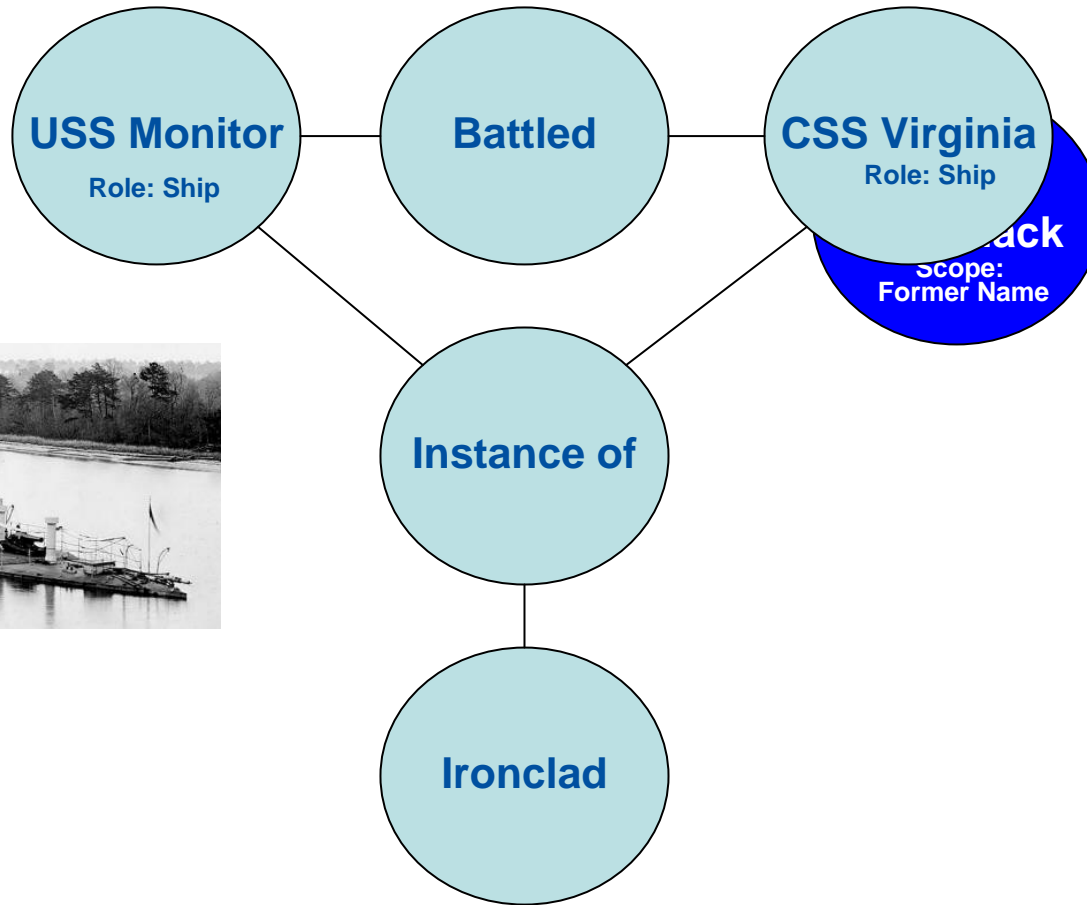
- The XML Topic Map ISO standard provides a standard, electronic means to express an ontology
- People can express their knowledge electronically in terms of topics and assertions about the topics, as well as links to external information sources
 - Web pages
 - Word documents
 - PDF files
 - Images
 - Etc.

Expertise Captured in Ontologies



- Because experts maintain the knowledge model, we can rely on that model to provide most users with the best results
- Our first attempt to connect the user with the information they desire is through the knowledge model

Modeling Knowledge



Challenges to Maintaining Ontologies



- Information may outstrip the ability of organizations to categorize
- The latest information may take time to be manually categorized
- One expert often cannot be completely familiar with the full breadth of information and terminology

An Enterprise Search Strategy



- A complete strategy for search will involve multiple techniques
- Search engines can provide powerful automatic categorization
- Ontological models can provide knowledge of experts
- Users should not have to make this choice actively, but instead should experience the best of both approaches

Unified Ontological and Traditional Search

monitor

Search

[USS Monitor](#)
[US Civil War](#)
[Ironclad](#)
[Cornelius Bushnell \(Shipwright\)](#)
[Gideon Welles \(Union Secretary\)](#)

- Knowledge model results are likely to be the most relevant to users
- As the knowledge model evolves, more and more useful information appears here – organized and categorized by experts

Documents related to your search

[The USS Monitor Center | The Mariners' Museum](#)

Site deals with the discovery and recovery of the Ironclad, plus its conservation, history, museum, and online resources.
www.monitorcenter.org/ - 21k

[Monitor Calibration](#)

Are things looking murky? Monitor calibration.
enr.com/news/energy/monitorcalibration

[Los Angeles International Airport](#)

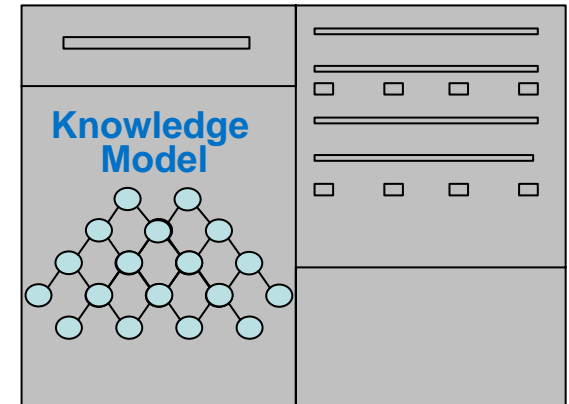
January, February, March, April, May, June, July, August, September, October, November, December. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31.
www4.passur.com/lax.html - 10k - [Cached](#) - [Similar pages](#)

- Search engine results are more comprehensive, but also more difficult to navigate
- The search engine manufacturer's features can be used as needed to organize the results
- Some engines can interact with the knowledge model or use their own to better organize this section

Further Integrating Ontologies with Engines

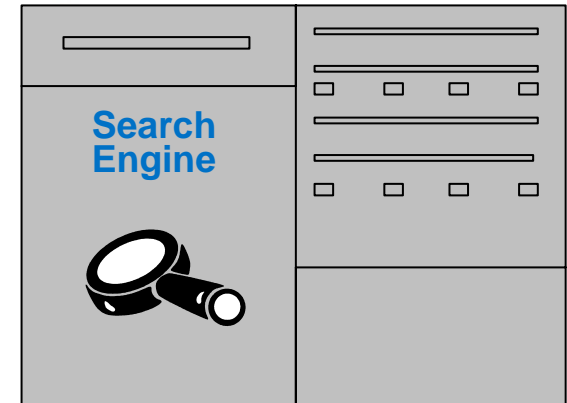
- Often the exact wording used in a search box does not match the words used in a topic map
 - Ontologies do support multiple names for a concept, but these cannot always be fully anticipated
- If research of text searches indicates some phrase being used instead of another, it can be easily added
- If no knowledge model results are found, this approach uses the search engine to attempt to fill in the gap

Searching Outside the Ontology



- The search for “Pook Turtles” does not produce results from the knowledge model

Traditional Search Engine Results are Obtained



[Pook Turtles](#) Pook Turtles "Backbone Of The River Fleet". "Pook Turtles" was the nickname given to seven ironclad Union gunboats designed by Samuel M. Pook to operate on ...

[Ironclads and Blockade Runners](#) The Pook Turtle Project - Building a replica of a western rivers ironclad gunboat: Union Naval Historical Society / Western Gunboat Flotilla · The Kearsarge ...

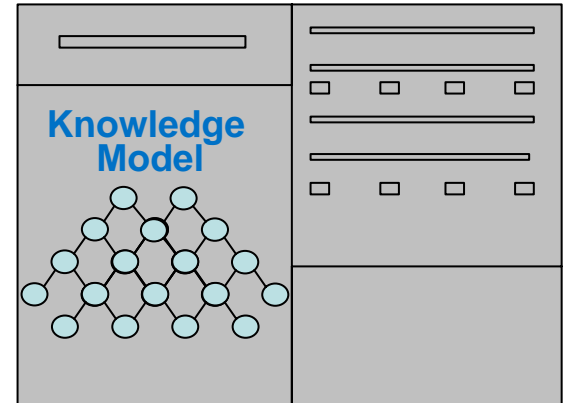
[Naval Living History Links](#) The Pook Turtle Project · Port Columbus Civil War Naval Museum · Naval Historical Center · Hampton Roads Naval Museum · Naval & Maritime Museums (Univ of ...

[Civil War Warship Innovation](#) "Pook Turtles" was the nickname given to seven Cairo class, also known as ... The Cairo, a "Pook Turtle," was the first warship in the world to be sunk by a ...

Using Documents to Query the Ontology



Pook Turtles Pook Turtles "Backbone Of The River Fleet". "Pook Turtles" was the nickname given to seven ironclad Union gunboats designed by Samuel M. Pook to operate on ...



USS Monitor

- The top documents found by the search engine are used to “reverse lookup” topics in the knowledge model
- These documents identify topics that users could have used to get “expert results” from the ontology

Enhanced Results

- Although the user searched for a term not in the Ontology (“Pook Turtles”), the combined approach used the search engine to map that term to a term in the ontology (“USS Monitor”)
- Users receive “expert” results combined with search engine results

[USS Monitor](#)
[US Civil War](#)
[Ironclad](#)
[Cornelius Bushnell \(Shipwright\)](#)
[Gideon Welles \(Union Secretary of the Navy\)](#)

Documents related to your search

[The USS Monitor Center | The Mariners' Museum](#)
Site deals with the discovery and recovery of the Ironclad, plus its conservation, history, museum, and online shop.
[www.monitorcenter.org/](#) - 21k - Oct 31, 2005 - [Cached](#) - [Similar pages](#)

Benefits of a Multi-Pronged Search



- Knowledge owners can express their expertise in connecting users to the proper information in an open format
- The knowledge model can be continually evolve in response to new ideas, approaches, and customer behaviors
 - The application can be launched with a modest knowledge model containing the most obvious topics and documents while the search engine fills in the gaps
 - As the knowledge model matures the user experience will improve

