

# Awareness Watch™ Newsletter

By Marcus P. Zillman, M.S., A.M.H.A.

<http://www.AwarenessWatch.com/>

V5N1

January 2007

Welcome to the V5N1 January 2007 issue of the **Awareness Watch™ Newsletter**. This newsletter is available as a complimentary subscription and will be issued monthly. Each newsletter will feature the following:

**Awareness Watch™ Featured Report**  
**Awareness Watch™ Spotters**  
**Awareness Watch™ Book/Paper/Article Review**  
**Subject Tracer™ Information Blogs**

I am always open to feedback from readers so please feel free to email with all suggestions, reviews and new resources that you feel would be appropriate for inclusion in an upcoming issue of Awareness Watch™. This is an ongoing work of creativity and you will be observing constant changes, constant updates knowing that “change” is the only thing that will remain constant!!

## **Awareness Watch™ Special Featured Report**

This month's featured report is a special report and I will not have the normal Awareness Watch™ Spotters, Awareness Watch™ Book/Paper/Article Review or related Subject Tracer™ Information Blogs. This special report will be on my latest research and the associated report/article that has been placed on line through LLRX.com . The title of my report is **Deep Web Research 2007** and is available at the following URLs:

<http://www.llrx.com/features/deepweb2007.htm>

<http://www.deepwebresearch.info/>

This comprehensive guide comprises hundreds of reliable sources to assist you in identifying best practices, strategies, websites and new technology applications which can be leveraged to mine the vast content of the deep web.



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

# Deep Web Research 2007

By

Marcus P. Zillman, M.S., A.M.H.A.

Bots, Blogs and News Aggregators is a keynote presentation that I have been delivering over the last several years, and much of my information comes from the extensive research that I have completed over the years into the “invisible” or what I like to call the “deep” web. The Deep Web covers somewhere in the vicinity of 900 billion pages of information located through the world wide web in various files and formats that the current search engines on the Internet either cannot find or have difficulty accessing. The current search engines find about 20 billion pages at the present time of this writing.

In the last several years, some of the more comprehensive search engines have written algorithms to search the deeper portions of the world wide web by attempting to find files such as .pdf, .doc, .xls, ppt, .ps. and others. These files are predominately used by businesses to communicate their information within their organization or to disseminate information to the external world from their organization. Searching for this information using deeper search techniques and the latest algorithms allows researchers to obtain a vast amount of corporate information that was previously unavailable or inaccessible. Research has also shown that even deeper information can be obtained from these files by searching and accessing the “properties” information on these files! This is interesting research that was written and posted in my personal blog a few months ago.

This article and guide is designed to give you the resources you need to better understand the history of the deep web research, as well as various classified resources that allow you to search through the currently available web to find those key sources of information nuggets only found by understanding how to search the “deep web”. This **Deep Web Research 2007** article is divided into the following sections:

- Articles, Papers, Forums, Audios and Videos
- Cross Database Articles
- Cross Database Search Services
- Cross Database Search Tools
- Peer to Peer, File Sharing, Grid/Matrix Search Engines
- Presentations
- Resources - Deep Web Research
- Resources - Semantic Web Research
- Bot Research Resources and Sites
- Subject Tracer Information Blogs



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**ARTICLES, PAPERS, FORUMS, AUDIOS AND VIDEOS (Current and Historical):**

**Academic and Scholar Search Engines and Sources**

<http://zillman.blogspot.com/2004/12/academic-and-scholar-search-engines.html>

**A Crisis for Web Preservation by Florence Olsen**

<http://snipurl.com/78te>

**All of OCLC's WorldCat Heading Toward the Open Web by Barbara Quint**

<http://www.infotoday.com/newsbreaks/nb041011-2.shtml>

**An Interactive Clustering-based Approach to Integrating Source Query interfaces on the Deep Web by W. Wu, C. Yu, A. Doan, W. Meng**

<http://www.cs.binghamton.edu/~meng/pub.d/sigmod04-final.pdf>

**Annotation for the Deep Web**

<http://csdl.computer.org/comp/mags/ex/2003/05/x5042abs.htm>

**Automatic Extraction of Web Search Interfaces for Interface Schema Integration by H. He, W. Meng, C. Yu, Z. Wu**

<http://www.cs.binghamton.edu/~meng/pub.d/WWWposterhe.pdf>

**Automatic Information Extraction From Semi-Structured Web Pages By Pattern Discovery**

<http://portal.acm.org/citation.cfm?id=640423&dl=ACM&amp;amp;coll=portal>

**Automatic Meaning Discovery Using Google by Rudi Cilibrasi and Paul M. B. Vitanyi**

<http://arxiv.org/abs/cs.CL/0412098>

**Benevolent "Virus" Helps Reveal the Hidden Web**

<http://www.syllabus.com/article.asp?id=9680>

**Beyond Google: The Invisible Web - Tools for Teaching the Invisible Web**

<http://www.lagcc.cuny.edu/library/invisibleweb/teachingtools.htm>

**Bibliomining Bibliography**

<http://biblio.syr.edu/bibliomining/articles/>



**Bibliomining for Automated Collection Development in a Digital Library Setting:  
Using Data Mining to Discover Web-Based Scholarly Research Works by Dr. Scott  
Nicholson**

<http://dlist.sir.arizona.edu/archive/00000625/>  
<http://www.BiblioMining.com/>

**Bot Research**

<http://www.BotResearch.info/>

**Client-Side Deep Web Data Extraction**

<http://doi.ieeecomputersociety.org/10.1109/CEC-EAST.2004.30>

**Clustering E-Commerce Search Engines by Q. Peng, W. Meng, H. He, C. Yu**

<http://www.cs.binghamton.edu/~meng/pub.d/WWWposterPeng.pdf>

**Common Information Environment Seeks To Reveal the Hidden Web**

<http://society.guardian.co.uk/e-public/story/0,13927,1195901,00.html>

**Crawling the Hidden Web by Sriram Raghavan and Hector Garcia-Molina**

<http://citeseer.ist.psu.edu/461253.html>

**Current Awareness Discovery Tools on the Internet**

<http://zillman.blogspot.com/2004/09/current-awareness-discovery-tools-on.html>

**Data Extraction and Label Assignment for Web Databases**

<http://www2003.sztaki.hu/cdrom/papers/refereed/p470/p470-wang.htm>

**Deep Content - Guide To Effective Searching of the Internet**

<http://www.brightplanet.com/deepcontent/tutorials/search/index.asp>

**Deep Web - Exploring the Secrets of the Hidden Internet by Marcus P. Zillman,  
M.S., A.M.H.A., - 23 minutes - Internet/Technology Channel**

<http://www.planetearthradio.com/technology.htm>

**Deep Web Navigation in Web Data Extraction**

<http://snipurl.com/13xdm>

**Desperately seeking Web Search 2.0**

<http://snipurl.com/64im>

**DigiCULT Thematic Issue 6**

**Resource Discovery Technologies for the Heritage Sector, June 2004**

Download Thematic Issue 6:Link HiRes .pdf (4,9 MB)

<http://snipurl.com/7v46>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Diving in the Deep End of the Web by Suzanne Ross**  
<http://research.microsoft.com/displayArticle.aspx?id=1052>

**Easy Topic Maps**  
<http://easytopicmaps.com/wiki/>

**Efficient and Effective Metasearch Project**  
<http://www.cs.binghamton.edu/~meng/metasearch.html>

**Farewell, Web 1.0! We Hardly Knew Ye by Steven Levy**  
<http://www.msnbc.msn.com/id/6214349/site/newsweek/>

**Fugitive Documents Evade Federal Depositories**  
<http://snipurl.com/78te>

**Google Teams Up with 17 Colleges to Test Searches of Scholarly Materials By Jeffrey R. Young**  
<http://chronicle.com/free/2004/04/2004040901n.htm>

**Graph Structure in the Web**  
<http://www9.org/w9cdrom/160/160.html>

**Gray Literature: Resources for Locating Unpublished Research by Brian S. Mathews**  
<http://snipurl.com/5i3b>

**Gray Literature Subject Guide**  
[http://www.csulb.edu/library/subj/gray\\_literature/](http://www.csulb.edu/library/subj/gray_literature/)

**Guardian Unlimited: Search for the Invisible Web**  
<http://www.guardian.co.uk/online/story/0,3605,547140,00.html>

**Indexing Deep Web Content By Paul Bruemmer**  
[http://www.searchengineguide.com/wi/2002/0327\\_wi2.html](http://www.searchengineguide.com/wi/2002/0327_wi2.html)

**Information Detective - The Invisible Web Mini-Tutorial Streaming Video**  
<http://www.InformationDetective.com/>

**Information Foraging and Extraction Techniques for Internet-Based Literature and Data**  
[http://www.asceditor.usm.edu/ASC%202006%20CD/2006pro/2006/CEGE03\\_Hannon06\\_7900.htm](http://www.asceditor.usm.edu/ASC%202006%20CD/2006pro/2006/CEGE03_Hannon06_7900.htm)



**Information Retrieval and the Semantic Web** by Tim Finin, James Mayfield, Clay Fink, Anupam Joshi, and R. Scott Cost  
<http://ebiquity.umbc.edu/v2.1/paper/html/id/185/>

**In Search of the Deep Web**  
[http://archive.salon.com/tech/feature/2004/03/09/deep\\_web/index\\_np.html](http://archive.salon.com/tech/feature/2004/03/09/deep_web/index_np.html)

**Invisible Web Gets Deeper**  
<http://www.searchenginewatch.com/sereport/article.php/2162871>

**Invisible Web Revealed**  
<http://www.searchenginewatch.com/sereport/article.php/2167321>

**IR and IE on the Web - PhD and MSc Dissertations**  
<http://www.webir.org/phd.html>

**JEP: The Deep Web**  
<http://www.press.umich.edu/jep/07-01/bergman.html>

**Library Journal: Braking Through the Invisible Web**  
<http://snipurl.com/5tbb>

**LLRX: Book Review: The Invisible Web**  
<http://www.llrx.com/features/invisibleweb.htm>

**LLRX: Deep Web Research**  
<http://www.llrx.com/features/deepweb.htm>

**LLRX: Deep Web Research 2005**  
<http://www.llrx.com/features/deepweb2005.htm>

**LLRX: Deep Web Research 2006**  
<http://www.llrx.com/features/deepweb2006.htm>

**LLRX: Mining Deeper Into the Invisible Web**  
<http://www.llrx.com/features/mining.htm>

**LLRX: ResearchWire: Exposing the Invisible Web**  
<http://www.llrx.com/columns/exposing.htm>

**Metadata? Thesauri? Taxonomies? Topic Maps!** by Lars Marius Garshol  
<http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>



**Mining Newsgroups Using Networks Arising From Social Behavior**  
<http://www2003.sztaki.hu/cdrom/papers/refereed/p688/688-agrawal/index.html>

**Mining the Deep Web With Specialized Drills**  
<http://snipurl.com/5tbd>

**Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews**  
<http://www2003.sztaki.hu/cdrom/papers/refereed/p451/package/p451-dave.html>

**Mining Topic-Specific Concepts and Definitions on the Web**  
<http://www2003.sztaki.hu/cdrom/papers/refereed/p646/p646-liu-XHTML/p646-liu.html>

**Modelling and Mining of Network Information Systems Publications**  
<http://www.mathstat.dal.ca/~mominis/Publications.htm>

**Net Plan Builds in Search by Kimberly Patch**  
<http://snipurl.com/5kn0>

**New Profusion Site Offers Better View of Invisible Web**  
<http://www.searchenginewatch.com/sereport/article.php/2163591>

**Noisy Channels Models Provide Short Answers to FAQs**  
[http://www.economist.com/printedition/displayStory.cfm?Story\\_ID=3127462](http://www.economist.com/printedition/displayStory.cfm?Story_ID=3127462)

**Old Search Engine, the Library, Tries to Fit Into a Google World**  
<http://snipurl.com/78rr>

**Online or Invisible?**  
<http://citeseer.ist.psu.edu/online-nature01/>

**OntoMiner: Bootstrapping and Populating Ontologies From Domain Specific Web Sites**  
<http://www.public.asu.edu/~hdavulcu/VLDB-WS03.pdf>

**OpenIndex - Creating a Public Internet Index**  
<http://www.openindex.org/index.php>

**PhysicsWeb: The Physics of the Web**  
<http://physicsweb.org/article/world/14/7/09>

**Publications about Web Analysis, Web Search, Citation Indexing, Digital Libraries, Machine Learning, Neural Networks [Steve Lawrence, Google Labs]**  
<http://labs.google.com/people/lawrence/>



**QProber: Classifying and Searching "Hidden-Web" Text Databases**  
<http://qprober.cs.columbia.edu/>

**Research Beyond Google: 119 Authoritative, Invisible, and Comprehensive Resources**  
<http://oedb.org/library/college-basics/research-beyond-google>

**Researcher Retrain Thyself**  
<http://www.infoday.com/online/sep04/OnTheNet.shtml>

**Researchers Map of the Web**  
[http://www.almaden.ibm.com/almaden/webmap\\_press.html](http://www.almaden.ibm.com/almaden/webmap_press.html)

**Scientific American: Featured Article: The Semantic Web**  
<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>

**Scraping the Web for Implied Data**  
<http://searchenginewatch.com/searchday/article.php/3374821>

**Search Engine Hunts for Gold Beneath the Surface of the Web**  
<http://snipurl.com/5tbe>

**Search Engine Meeting 2005 Boston, Massachusetts - White Papers and Presentations**  
<http://www.infonortics.com/searchengines/sh05/05pro.html>

**Search Engine Technology and Digital Libraries**  
<http://www.dlib.org/dlib/june04/lossau/06lossau.html>

**Searching the Deep Web**  
<http://www.dlib.org/dlib/january01/warnick/01warnick.html>

**Searching the Deep Web - Video**  
<http://www.osti.gov/media/DeepWebVideo.html>

**Searching the Internet (White Paper, Audio and Video)**  
<http://www.SearchingTheInternet.info/>

**Seeing through the 'invisible' Web**  
<http://www.usatoday.com/tech/2001/10/15/invisible-web-search.htm>





**Semantic Web Content Accessibility Guidelines for Current Research Information Systems (CRIS) by A. Lopatenko**

[http://derpi.tuwien.ac.at/~andrei/AURIS\\_DE.htm](http://derpi.tuwien.ac.at/~andrei/AURIS_DE.htm)

**Smart Search - Advanced Search Engines Link Many Data Sources**

[http://gcn.com/23\\_24/tech-report/26999-1.html](http://gcn.com/23_24/tech-report/26999-1.html)

**Structured Databases on the Web: Observations and Implications**

<http://eagle.cs.uiuc.edu/pubs/2004/dwsurvey-sigmoidrecord-chlpz-aug04.pdf>

**Testbed for Information Extraction from Deep Web**

<http://www2004.org/proceedings/docs/2p346.pdf>

**The Deep Web**

<http://www.internettutorials.net/deepweb.html>

**The Deep Web: Surfacing Hidden Value by Michael K. Bergman**

<http://www.press.umich.edu/jep/07-01/bergman.html>

**The Future Of News: The Digital Information Librarian**

[http://www.masternewmedia.org/2004/03/24/the\\_future\\_of\\_news\\_the.htm](http://www.masternewmedia.org/2004/03/24/the_future_of_news_the.htm)

**The Hidden Potential of the Web**

<http://society.guardian.co.uk/e-public/story/0,13927,1195901,00.html>

**The Invisible Web by Chris Sherman**

<http://www.freepint.com/issues/080600.htm#feature>

**The Invisible Web: What it is, Why it exists, How to find it, and Its Inherent Ambiguity**

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>

**The Invisible Web: Where Search Engines Fear To Go**

<http://www.powerhomebiz.com/vol25/invisible.htm>

**The Mechanics of Deep Net Meta Search**

<http://turbo10.com/papers/deepnet.pdf>

**The Seventh Asia Pacific Web Conference (APWeb05)**

<http://apweb05.csm.vu.edu.au/index.asp>

**Topological Measures and Maps Of the Web**

<http://informatics.indiana.edu/fil/Web/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Towards Automatic Incorporation of Search Engines Into A Large-Scale Metasearch Engine**

<http://www.cs.binghamton.edu/~meng/pub.d/wi2003.pdf>

**Traffic-Based Feedback on the Web by Jonathan Aizen, Daniel Huttenlocher, Jon Kleinberg, and Antal Novak**

<http://www.pnas.org/cgi/content/abstract/0307539100v1>

**UMBC - AgentNews**

<http://agents.umbc.edu/agentnews/>

**Understanding Metadata**

<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

**Using the Internet As a Dynamic Resource Tool for Knowledge Discovery**

<http://zillman.blogspot.com/2004/09/using-internet-as-dynamic-resource.html>

**Web Characterization Project**

<http://wcp.oclc.org/>

**Web Data Extractors White Paper Link Compilation**

<http://zillman.blogspot.com/2004/09/web-data-extractors.html>

**Web Pages Search Engine Based on DNS by Wang Liang, Guo Yi-Ping, and Fang Ming**

<http://arxiv.org/pdf/cs.NI/0403035>

**WebScales: Towards a Highly Scalable Metasearch Engine**

<http://www.cs.binghamton.edu/~meng/pub.d/Pireport04.html>

**What Is the Deep Web? A WhatIs Podcast 15 Minute Interview with Marcus P. Zillman**

<http://zillman.blogspot.com/2006/10/what-is-deep-web.html>

**What is the Invisible Web? A Crawler Perspective by Natalia Arroyo, Laboratorio de Internet**

<http://cybermetrics.wlv.ac.uk/AoIRASIST/arroyo.html>

**WISE-Cluster: Clustering E-Commerce Search Engines Automatically by Q. Peng, W. Meng, H. He, C. Yu**

<http://www.cs.binghamton.edu/~meng/pub.d/PengWIDM04.pdf>

**Yahoo and the Deep Web**

<http://news.com.com/2100-1024-5167931.html>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**ZDNet: I've Discovered the 'invisible Web'--Have You? Here's How!**  
[http://reviews-zdnet.com.com/4520-6033\\_16-4206148.html](http://reviews-zdnet.com.com/4520-6033_16-4206148.html)

**CROSS DATABASE ARTICLES:**

**Digital Libraries- Cross-Database Search: One-Stop Shopping**  
<http://snipurl.com/5tbf>

**Search Tools Reports: Searching for Text Information in Databases**  
<http://www.searchtools.com/info/database-search.html>

**The Right Solution: Federated Search Tools by Roy Tennant**  
<http://snipurl.com/5zxp>

**UK Web Archiving Consortium**  
<http://www.webarchive.org.uk/>

**CROSS DATABASE SEARCH SERVICES:**

**Entrez - The Life Sciences Cross-Database Search Engine**  
<http://www.ncbi.nlm.nih.gov/Entrez/index.html>

**EnergyFiles - Subject Pathways**  
<http://energyfiles.osti.gov/>

**GPO Access - Search Across Multiple Databases**  
<http://www.gpoaccess.gov/multidb.html>

**King County Library System**  
<http://www.kcls.org/>

**NLM Gateway Search**  
<http://gateway.nlm.nih.gov/gw/Cmd>

**SUMSearch**  
<http://sumsearch.uthscsa.edu/>

**The Metasearch Infrastructure Project**  
<http://www.cdlib.org/inside/projects/metasearch/>



## CROSS DATABASE SEARCH TOOLS

### Apple - Mac - Sherlock

<http://www.asia.apple.com/sherlock/>

### Blue Angel Technologies

<http://www.blueangeltech.com/>

### Bright Planet

<http://brightplanet.com/>

### Copernic

<http://www.copernic.com/en/index.html>

### Cross Database Search Tools Summary

<http://lists.webjunction.org/wjlists/web4lib/2001-September/027669.html>

### DbVisualizer - The Universal Database Tool

<http://www.dbvis.com/products/dbvis/>

### Dublin Core Metadata Initiative (DCMI)

<http://www.dublincore.org/>

### EEVL Xtra - Cross Database Search

<http://www.eevlextra.ac.uk/>

### EMC

<http://software.emc.com/>

### Gold Rush - Database Search Tool

<http://goldrush.coalliance.org/>

### MetaLib

<http://www.exlibrisgroup.com/metilib.htm>

### MetaSearch Initiative

[http://www.niso.org/committees/MS\\_initiative.html](http://www.niso.org/committees/MS_initiative.html)

### mod\_oai Project - Getting OAI-PMH For Free

<http://www.modoi.org/>

### MuseGlobal

<http://www.museglobal.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Peter's PolySearch Engines**

<http://www2.hawaii.edu/~jacso/extra/poly-page.html>

**PBCore - The Public Broadcasting Metadata Dictionary**

<http://www.utah.edu/cpbmetadata/>

**Registry of Library Knowledge Bases**

<http://www.public.iastate.edu/~CYBERSTACKS/KBL.htm>

**Search Federal Research and Development**

<http://fedrnd.osti.gov/>

**SRU - Search/Retrieve via URL**

<http://www.loc.gov/standards/sru>

**STINET Multisearch**

<http://multisearch.dtic.mil/>

**The Flamenco Search Interface Project**

<http://bailando.sims.berkeley.edu/flamenco.html>

**VIAF: The Virtual International Authority File**

<http://www.oclc.org/research/projects/viaf/default.htm>

**WebFeat**

<http://www.webfeat.org/>

**PEER TO PEER (P2P), FILE SHARING, GRID AND MARIX SEARCH ENGINES:**

**AllPeers**

<http://www.allpeers.com/>

**ALPINE Network - SourceForge: Project**

<http://sourceforge.net/projects/alpine/>

**An Efficient Scheme for Query Processing on Peer-to-Peer Networks**

<http://aeolusres.homestead.com/files/index.html>

**angrycoffee.com**

<http://www.AngryCoffee.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Azureus - Java Bittorrent Client**

<http://azureus.sourceforge.net/>

**BadBlue**

<http://badblue.com/>

**Between Rhizomes and Trees: P2P Information Systems by Bryn Loban**

[http://www.firstmonday.org/issues/issue9\\_10/loban/index.html](http://www.firstmonday.org/issues/issue9_10/loban/index.html)

**Bibster**

<http://bibster.semanticweb.org/index.htm>

**BigChampagne**

<http://www.bigchampagne.com/>

**BitTorrent FAQ and Guide**

<http://www.dessent.net/btfaq/>

**Bit Torrent Official Site and Search Engine**

<http://www.BitTorrent.com/>

**Bitzi - The Free Universal Media Catalog**

<http://www.bitzi.com/>

**Blog Torrent**

<http://www.blogtorrent.com/>

**Blubster**

<http://www.blubster.com/>

**BotSpot®: File-sharing Bots**

[http://www.botspot.com/BOTSPOT/Windows/Download\\_Bots/File-sharing\\_Bots/](http://www.botspot.com/BOTSPOT/Windows/Download_Bots/File-sharing_Bots/)

**BTbot - BitTorrent Search Engine**

<http://www.btbot.com/>

**Coral - The Coral P2P Content Distribution Network**

<http://www.coralcdn.org/>

**Capn's PHP Gnutella Search**

<http://capnbry.net/gnutella/gs.php>

**Current P2P Search Implementations - P2P Networks**

<http://ntrg.cs.tcd.ie/undergrad/4ba2.02-03/p8.html#CurrentP2PSearchImplementations>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**DebateRoom.com - XDCC Search / File Sharing Portal**

<http://www.debateroom.com/>

**Deepnet Explorer - P2P/RSS-ATOM Web Browser**

<http://www.deepnetexplorer.com/>

**Distributed Search Engines**

<http://www.openp2p.com/pub/t/74>

**Distributed Search in P2P Networks**

<http://csdl.computer.org/comp/mags/ic/2002/01/w1068abs.htm>

**Filetopia**

<http://www.filetopia.org/>

**Free Haven Project**

<http://www.freehaven.net/index.html>

**FuzzBox: Tangent Research Artificial Intelligence and Robotics**

<http://tangentresearch.com/research/ai/>

**Gnougat: Fully decentralised file caching from the JXTA Project**

<http://gnougat.jxta.org/>

**GNUnet - GNU Project - Free Software Foundation (FSF)**

<http://www.gnu.org/software/GNUnet/gnunet.html>

**Gnutella.com**

<http://www.gnutella.com/>

**gPulp**

<http://www.gpulp.com/>

**GRACE IST Project**

<http://www.grace-ist.org/>

**GRACE - GRid seArch and Categorization Engine**

<http://www.ub.uni-stuttgart.de/grace/>

**Grid Resources**

<http://www.GridResources.info/>

**Grokster3G**

<http://www.grokster3g.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Grouper - P2P Personal Media File Sharing**

<http://www.grouper.com/>

**grub.org - Open Source, Distributed Internet Crawler!**

<http://grub.org/>

**Hamachi - Secure Mediated Peer To Peer**

<http://www.hamachi.cc/>

**HyperCuP – Shaping Up Peer-to-Peer Networks**

<http://www-db.stanford.edu/~schloss/hypercup/>

**Ian Clarke's Blog**

<http://locut.us/blog/index.php>

**IM and P2P Threat Center**

[http://www.imlogic.com/im\\_threat\\_center/index.asp](http://www.imlogic.com/im_threat_center/index.asp)

**iMesh**

<http://www.iMesh.com/>

**International Workshop on Peer-to-Peer Knowledge Management (P2PKM)**

<http://www.p2pkm.org/>

**Internet Movie Database (IMDb)**

<http://www.imdb.com/>

**isoHunt - IRC and Bit Torrent Search Engine**

<http://isohunt.com/>

**JXTA Project**

<http://www.jxta.org/>

**Kademlia: A Peer-to-peer Information System Based on the XOR Metric**

<http://citeseer.ist.psu.edu/529075.html>

**Kazaa Media Desktop**

<http://www.kazaa.com/us/index.htm>

**Legal P2P File Sharing Software**

<http://www.filesharesoftware.com/>

**LegalTorrents**

<http://www.legaltorrents.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.



**Limewire**

<http://www.limewire.com/>

**LionShare P2P Project - Legitimate File-Sharing Among Individuals and Educational Institutions**

<http://lionshare.its.psu.edu>

**MagnetLink**

<http://www.magnetlinks.org/>

**Mercora IM P2P Radio**

<http://www.mercora.com/>

**MoleSter - A Tiny File-Sharing Application**

<http://ansuz.sooke.bc.ca/software/molester/>

**Mnet**

<http://mnet.sourceforge.net/>

**Morpheus :: Peer-to-Peer File Sharing Software**

<http://www.morpheus.com/>

**MusicBrainZ**

<http://www.MusicBrainZ.org/>

**MysterNetworks - The Evolution of Peer-to-Peer**

<http://www.mysternetworks.com/>

**NeuroGrid - P2P Search**

<http://www.neurogrid.net/>

**Open Directory - File Sharing**

[http://dmoz.org/Computers/Software/Internet/Clients/File\\_Sharing/](http://dmoz.org/Computers/Software/Internet/Clients/File_Sharing/)

**Open Directory - MP3 Search Engines**

[http://dmoz.org/Arts/Music/Sound\\_Files/MP3/Search\\_Engines/](http://dmoz.org/Arts/Music/Sound_Files/MP3/Search_Engines/)

**OpenNap: Open Source Napster Server**

<http://opennap.sourceforge.net/>

**OpenP2P.com**

<http://www.openp2p.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Oyster - Managing, Searching and Sharing Ontology Metadata in a Peer-to-Peer Network.**

<http://oyster.ontoware.org/>

**P2P and the Future of Private Copying by Peter K. Yu, Michigan State University College of Law**

[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=578568](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=578568)

**P2PNet - Updated P2P News**

<http://p2pnet.net/index.php>

**P2P News from Topex**

<http://www.topix.net/tech/p2p>

**PeerCast P2P Radio**

<http://www.peercast.org/>

**PeerMetrics.org**

<http://www.peermetrics.org/>

**PeerMind - P2P Monitor**

<http://www.PeerMind.com/>

**Piolet**

<http://www.piolet.com/>

**Port Knocking**

<http://www.portknocking.org/>

**PowerFolder - P2P Whole Folder Synchronization**

<http://www.powerfolder.com/>

**Project JXTA**

<http://www.jxta.org/>

**Rodi - Tiny P2P Client/Host**

<http://larytet.sourceforge.net/btRat.shtml>

**ScrapeTorrent**

<http://www.ScrapeTorrent.com/>

**Shareaza**

<http://www.shareaza.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**ShareSniffer**

<http://www.sharesniffer.com/>

**Skype**

<http://www.skype.com/>

**Slyck - File Sharing News and Info**

<http://www.slyck.com/index.php>

**Snoopstar**

<http://www.snoopstar.com/>

**Streamload - Share Videos and Photos - Online MP3 Storage and Access**

<http://www.streamload.com/>

**Super-Peer-Based Routing and Clustering Strategies for RDF-Based Peer-to-Peer Networks**

<http://citeseer.ist.psu.edu/nejdl02superpeerbased.html>

**Super Powered Peer To Peer**

<http://snipurl.com/9lzg>

**SwarmStream™ SDK**

<http://onionnetworks.com/products/swarmstream/>

**The Anthill Project**

<http://www.cs.unibo.it/projects/anthill/>

**The Pirate Bay - BitTorrent Tracker**

<http://thepiratebay.org/>

**The Chord Project**

<http://pdos.csail.mit.edu/chord/>

**The Freenet Project**

<http://freenetproject.org/>

**The Peer-to-Peer Weblog**

<http://p2p.weblogsinc.com/>

**The Role of Peer to Peer File Sharing in Law Firm Marketing by Andy Havens**

<http://www.llrx.com/columns/marketing7.htm>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**ToPeer**

<http://www.topeer.com/>

**Torrent Finder**

<http://ts.kurtubba.com/>

**Torrent Reactor**

<http://www.torrentreactor.net/>

**Torrent Typhoon (TT)**

<http://www.torrenttyphoon.com/>

**TrustyFiles**

<http://www.trustyfiles.com/>

**Understanding BitTorrent: An Experimental Perspective by Arnaud Legout, Guillaume Urvoy-Keller, and Pietro Michiardi**

<http://hal.inria.fr/inria-00000156/en>

**URLBlaze: URL Sharing Network**

<http://www.urlblaze.com/>

**Videora - Personal Video Using P2P and RSS**

<http://www.videora.com/>

**WASTE**

<http://slackerbitch.free.fr/waste/>

**WiredReach - Powering the User Centric Web**

<http://www.wiredreach.org/>

**Yahoo! Directory Peer-to-Peer File Sharing**

[http://dir.yahoo.com/Computers\\_and\\_Internet/Internet/Peer\\_to\\_Peer\\_File\\_Sharing/](http://dir.yahoo.com/Computers_and_Internet/Internet/Peer_to_Peer_File_Sharing/)

**YAPPERS: A Peer-to-Peer Lookup Service over Arbitrary Topology**

<http://citeseer.ist.psu.edu/ganesan03yappers.html>

**YouServ - A P2P (peer-to-peer) Web Hosting/File Sharing System**

<http://www.bayardo.org/youserv/>

**Zebra**

<http://indexdata.dk/zebra/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

## **PRESENTATIONS:**

### **From Theory To Practice - Bielefeld Academic Search Engine**

<http://www.diglib.org/forums/Spring2004/summann0404.htm>

### **Gumshoe Librarian**

<http://www.llrx.com/features/gumshoe.htm>

### **Information Detective - Online Streaming Tutorial Videos On Searching the Internet including the Deep and Invisible Web**

<http://www.InformationDetective.com/>

### **Quick Introduction to OWL Web Ontology Language**

<http://www.iro.umontreal.ca/~lapalme/ift6281/OWL/CostelloQuickIntroOwl.pdf>

### **Searching the Deep Web - Dudley Knox Library Internet Guides - PowerPoint Slides**

<http://library.nps.navy.mil/home/Searching%20the%20Deep%20Web.ppt>

### **Searching the Internet**

<http://www.SearchingTheInternet.info/>

### **Searching the Internet: Using Brains and Bots**

<http://snipurl.com/5kza>

### **Seeing the Invisible Web**

<http://lib.berkeley.edu/TeachingLib/Guides/Internet/InvWebPowerpoint/index.htm>

## **RESOURCES – DEEP WEB RESEARCH:**

### **A Roadmap for Web Mining: From Web to Semantic Web**

<http://eprints.pascal-network.org/archive/00000841/01/roadmap.pdf>

### **Beaucoup**

<http://www.beaucoup.com/>

### **BlogPulse**

<http://www.BlogPulse.com/>

### **Bot Research**

<http://www.BotResearch.info/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**BrainBoost - Question Answering Search Engine**

<http://www.BrainBoost.com/>

**BrightPlanet's Deep Federation Portal™ (DFP)**

<http://www.brightplanet.com/products/dfportal.asp>

**Can't Find On Google**

<http://www.cantfindongoogle.com>

**CiteLine Professional**

[http://www.citeline.com/pro\\_info.html](http://www.citeline.com/pro_info.html)

**COLLATE - Collaboratory for Annotation, Indexing and Retrieval of Digitized Historical Archive Material**

<http://www.collate.de/>

**Comet Way**

[http://www.cometway.com/content.agent?page\\_name=Home](http://www.cometway.com/content.agent?page_name=Home)

**CompletePlanet - 70,000 Databases and Speciality Search Engines**

<http://www.completeplanet.com/>

**Creative Commons RDF-Enhanced Search**

<http://search.creativecommons.org/>

**Cyber Cemetery**

<http://govinfo.library.unt.edu/>

**CyberFiber**

<http://www.cyberfiber.com>

**Cybermetrics - First Generation Tools - Invisible Web**

<http://www.cindoc.csic.es/cybermetrics/search13.html>

**Data Fountains: Open Source Internet Resource Discovery and Metadata/Full-Text Generation Service**

[http://infomine.ucr.edu/Data\\_Fountains/](http://infomine.ucr.edu/Data_Fountains/)

**Data Mining Resources**

<http://www.DataMiningResources.info/>

**Deep Web Research**

<http://www.DeepWebResearch.info/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Deep Web Search**

<http://www.mach9design.com/deep/deep1.html>

**Deep Web Technologies**

<http://www.deepwebtech.com/>

**DigiCULT Resources - Resource Discovery & Information Retrieval**

<http://www.digicult.info/pages/resources.php?t=21>

**digitalAGORA**

<http://aut.edu/agora/>

**Direct Search**

<http://www.freepint.com/gary/direct.htm>

**EEVL's Ejournal Search Engines**

<http://www.eevl.ac.uk/eese/eese-eevl.html>

**ENDECA**

<http://www.endeca.com/>

**Engineering Village 2**

<http://www.engineeringvillage2.org/>

**Find Articles**

<http://www.findarticles.com/PI/index.jhtml>

**Freely Accessible Databases for the Public**

<http://www.istl.org/01-winter/internet.html>

**Ghostscript, Ghostview and GSview**

<http://www.cs.wisc.edu/~ghost/>

**GlobalSpec - Engineering Search Engine**

<http://search.globalspec.com/Search/WebSearch>

**Google Labs**

<http://labs.google.com/>

**Google Scholar**

<http://scholar.google.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**HighWire Press - Largest Repository of Free Full-Text Life Science Articles in the World**

<http://highwire.stanford.edu/>

**iBoogie™**

<http://www.iboogie.tv/>

**IncyWincy - The Invisible Web Search Engine**

<http://www.incywincy.com/>

**INFOMINE**

<http://infomine.ucr.edu/>

**Instant Information Systems**

<http://www.docdel.com/>

**Institutional Archives Registry**

<http://archives.eprints.org/eprints.php?action=browse>

**Intelligence Center**

<http://www.intelligence-center.com/>

**Intellisonar™**

<http://www.quigo.com/intellisonar.htm>

**Internet Archive**

<http://www.archive.org/>

**Intute**

<http://www.intute.ac.uk/>

**Invisible Library**

<http://sanchezkisser.com/blog/>

**Kapow Web Collector**

<http://www.automated-info-solutions.com/>

**KDnuggets: Data Mining, Web Mining, and Knowledge Discovery Guide**

<http://www.kdnuggets.com/>

**KeepMedia**

<http://www.keepmedia.com/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.



**Knowledge Discovery**

<http://www.KnowledgeDiscovery.info/>

**Large-Scale Deep Web Integration: Incomplete Bibliography**

<http://metaquerier.cs.uiuc.edu/webibib.html>

**Librarians' Index to the Internet**

<http://lii.org/>

**MagPortal**

<http://www.magportal.com/>

**Mamma - Deep Web Health Search Engine**

<http://www.mammahealth.com/>

**Mappa.Mundi Magazine**

<http://mappa.mundi.net/>

**Medical Databases Online**

<http://www.medic8.com/MedicalDatabases.htm>

**Microsoft Web Search Research and Patents**

<http://www.webmasterworld.com/forum97/5.htm>

**Mining the Deep Web for Economic Data**

[http://www.citris-uc.org/research/projects/mining\\_the\\_deep\\_web\\_for\\_economic\\_data](http://www.citris-uc.org/research/projects/mining_the_deep_web_for_economic_data)

**Mooter Search**

<http://www.mooter.com/>

**MSN Sandbox**

<http://sandbox.msn.com/>

**NetScan**

<http://netscan.research.microsoft.com/>

**News Group Search**

<http://newsgroups.langenberg.com/>

**New Zealand Digital Library**

<http://www.nzdl.org/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**OAI-PMH Implementation Guidelines - Conveying rights expressions about metadata in the OAI-PMH framework**

<http://www.openarchives.org/OAI/2.0/guidelines-rights.htm>

**OAIster**

<http://oaister.umdl.umich.edu/o/oaister/>

**OneLook Dictionary Search**

<http://www.onelook.com/>

**Open Archives Initiative**

<http://www.openarchives.org/>

**OpenIndex - Creating a Public Internet Index**

<http://www.openindex.org/index.php>

**Open WorldCat-enabled Web Tools**

<http://www.oclc.org/ca/fr/worldcat/open/searchtools/default.htm>

**QProber: Classifying and Searching "Hidden-Web" Text Databases - PERSIVAL Project**

<http://qprober.cs.columbia.edu/>

**Quigo Technologies**

<http://www.quigo.com/>

**Pretrieve Search - Free Public Record Search Engine**

<http://www.pretrieve.com/>

**Recommended Gateway Sites for the Deep Web**

<http://people.hws.edu/hunter/deepwebgate03.htm>

**reSearcher**

<http://researcher.sfu.ca/>

**Resource Discovery Network**

<http://www.rdn.ac.uk/>

**Science and Technology Sources on the Internet**

<http://www.library.ucsb.edu/istl/01-winter/internet.html>

**Scientific and Technical Information Network (STINET)**

<http://stinet.dtic.mil/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Science Commons**

<http://science.creativecommons.org/>

**Science.gov - FirstGov for Science - Government Science Portal**

<http://www.science.gov/>

**Scirus - Search Engine for Scientific Information**

<http://www.scirus.com/srsapp/>

**SDARTS - A Protocol and Toolkit for Metasearching**

<http://sdarts.cs.columbia.edu/>

**Search Adobe PDF Online**

<http://www.SearchPDF.com/>

**STN International - Databases in Science and Technology**

<http://www.stn-international.de/>

**TechXtra - Indepth Academic and Scholar Search**

<http://www.techxtra.ac.uk/>

**Testbed for Information Extraction from Deep Web**

[http://research.microsoft.com/users/nickcr/pubs/yamada\\_www2004poster.pdf](http://research.microsoft.com/users/nickcr/pubs/yamada_www2004poster.pdf)

**The Internet Sleuth**

<http://www.isleuth.com/>

**The Deep Web**

<http://www.internettutorials.net/deepweb.html>

**The Invisible Web**

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>

**THOR: Deep Web Data Extraction**

<http://disl.cc.gatech.edu/THOR/>

**Those Dark Hiding Places: The Invisible Web Revealed**

<http://www.robertlackie.com/invisible/index.html>

**Turbo10**

<http://turbo10.com/>

**UNESCO Information Services - Databases**

<http://www.unesco.org/unesdi/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Wall Street Executive Library**  
<http://www.executivelibrary.com/>

**Web Data Extractors**  
<http://zillman.blogspot.com/2004/09/web-data-extractors.html>

**Web Farming**  
<http://webfarming.com/>

**WebFountain™**  
<http://www.research.ibm.com/journal/sj/431/gruhl.html>

**Web Intelligence Consortium**  
<http://wi-consortium.org/>

**Web IR & IE**  
<http://www.webir.org/>

**WebScales: Towards a Highly Scalable Metasearch Engine**  
<http://www.cs.binghamton.edu/~meng/pub.d/PIreport04.html>

**Web-Searching Agents**  
<http://www.aaai.org/AITopics/html/webagent.html>

## **RESOURCES - SEMANTIC WEB RESEARCH:**

**AIS SIGSEMIS - SIGSEMIS: Semantic Web and Information Systems**  
<http://www.sigsemis.org/>

**Analyzing Social Networks on the Semantic Web**  
<http://snipurl.com/cbdq>

**Bibster**  
<http://bibster.semanticweb.org/index.htm>

**Combining RDF and OWL with SOAP for Semantic Web**  
<http://www.ida.liu.se/~yuxzh/doc/ncws-041002.pdf>

**Cypher - Plain Language Access to the Semantic Web**  
<http://www.monrai.com/products/cypher>



**DARPA Agent Markup Language**

<http://www.daml.org/>

**DBin Project - Semantic Web P2P and/or Semantic Newsgroup Client.**

<http://www.dbin.org/>

**DERI International - Digital Enterprise Research Institute**

<http://www.deri.org/>

**Digital Object Identifier (DOI)**

<http://www.doi.org/>

**Dublin Core Services**

<http://www.describethis.com/>

**Fabl - A Native Programming Language for the Semantic Web**

<http://fabl.net/>

**Foundation for Intelligent Physical Agents (FIPA)**

<http://www.fipa.org/>

**The FOAF Project - A Semantic Web Application**

<http://www.foaf-project.org/>

**HP Labs Semantic Web Research**

<http://www.hpl.hp.com/semweb/index.html>

**Infomesh's Semantic Web Introduction**

<http://infomesh.net/2001/swintro/>

**International Journal of Metadata, Semantics and Ontologies (IJMSO)**

<http://www.inderscience.com/browse/index.php?journalCODE=ijms>

**Jena – A Semantic Web Framework for Java**

<http://jena.sourceforge.net/>

**Journal of Web Semantics: Preprint Server**

<http://www.websemanticsjournal.org/>

**KnowledgeNets**

<http://www.inf.fu-berlin.de/inst/ag-nbi/research/wissensnetze/>

**Knowledge Search**

<http://www.KnowledgeSearch.org/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Language Engineering for the Semantic Web: A Digital Library for Endangered Languages**

<http://informationr.net/ir/9-3/paper176.html>

**Magpie - The Samatic Filter and Tool For the Semantic Web**

<http://kmi.open.ac.uk/projects/magpie/main.html>

**MetaData at W3C**

<http://www.w3.org/Metadata/>

**Metadata FAQ - Metadata for Education**

<http://www.cetis.ac.uk/metadafaq/FrontPage>

**MindRaider - Semantic Web Outliner**

<http://mindraider.sourceforge.net/>

**MindSwap**

<http://www.MindSwap.org/>

**MuseoSuomi**

<http://museosuomi.cs.helsinki.fi/>

**OASIS - Advancing eBusiness Standards**

<http://www.oasis-open.org/home/index.php>

**OIL - Ontology Inference Layer**

<http://www.ontoknowledge.org/oil/index.shtml>

**Ontologies for Education (O4E)**

<http://iiscs.wssu.edu/o4e/>

**Ontology Matching**

<http://www.ontologymatching.org/>

**Ontology Metadata Vocabulary (OMV)**

<http://omv.ontoware.org/>

**OntoWare**

<http://ontoware.org/>

**O'Reilly's Semantic Web Primer**

<http://www.xml.com/pub/a/2000/11/01/semanticweb/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Potential Advantages Of Semantic Web For Internet Commerce by Yuxiao Zhao and Kristian Sandahl**

<http://www.ida.liu.se/~yuxzh/doc/iceis-030120.pdf>

**pOWL - Semantic Web Development Plattform**

<http://powl.sourceforge.net/>

**Practical Semantic Analysis of Web Sites and Documents**

<http://www.www2004.org/proceedings/docs/1p685.pdf>

**RDF Context Tools**

<http://www.dbin.org/RDFContextTools.php>

**RDF - Resource Description Framework**

<http://www.w3.org/RDF/>

**RDFWeb: Friend of a Friend (FOAF) Project**

<http://rdfweb.org/>

**Rules and Rule Markup Languages for the Semantic Web - RuleML-2003**

<http://www.informatik.uni-trier.de/~ley/db/conf/semweb/ruleml2003.html>

**Science and the Semantic Web**

<http://www.mindswap.org/Science/>

**Semantic Blogging: Spreading the Semantic Web Meme**

[http://jena.hpl.hp.com/~stecay/papers/xml europe2004/040420\\_semblog\\_draft10.html](http://jena.hpl.hp.com/~stecay/papers/xml europe2004/040420_semblog_draft10.html)

**gnowsis - Semantic Desktop Environment**

<http://www.gnowsis.org/>

**Semantic Email by Luke McDowell, Oren Etzioni, Alon Halevy, and Henry Levy**

<http://www.cs.usna.edu/~lmcdowne/>

**Semantic Indexing**

<http://www.nitle.org/index.php/nitle/resources/semantic>

**Semantic Interoperability of Metadata and Information in unLike Environments (SIMILE)**

<http://simile.mit.edu/>

**Semantic Knowledge Technologies and Language Computation**

<http://gate.ac.uk/projects/sekt/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Semantic Markup Deconstructed Example**

<http://www.cs.umd.edu/users/hendler/sciam/walkthru.html>

**Semantic Planet Weblog**

<http://www.semanticplanet.com/>

**Semantic Routing BOF**

<http://www.neurogrid.net/SemanticRouting/SemanticRoutingBOF.htm>

**Semantic Translator for Enhanced Retrieval by the Bremen University (BUSTER)**

<http://www.informatik.uni-bremen.de/agki/www/buster/new/application.html>

**SemanticWeb.org - The Semantic Web Community Portal**

<http://www.semanticweb.org/>

**Semantic Web Activity Statement**

<http://www.w3.org/2001/sw/Activity.html>

**Semantic Web Application Platform - SWAP**

<http://www.w3.org/2000/10/swap/>

**Semantic Web for AURIS-MM**

<http://derpi.tuwien.ac.at/~andrei/AURIS-MM-plan.html>

**Semantic Web Laboratory**

[http://iit-iti.nrc-cnrc.gc.ca/projects-projets/sem-web-lab-web-sem\\_e.html](http://iit-iti.nrc-cnrc.gc.ca/projects-projets/sem-web-lab-web-sem_e.html)

**Semantic Web Primer for Object-Oriented Software Developers**

<http://www.w3.org/TR/2006/NOTE-sw-oosd-primer-20060309/>

<http://www.w3.org/2001/sw/>

**Semantic Web Publications**

<http://www.w3.org/2001/sw/#pub>

**Semantic Web Roadmap**

<http://www.w3.org/DesignIssues/Semantic.html>

**Semantic Web Services Challenge 2006**

<http://www.sws-challenge.org/>

**Semantic Web W3C**

<http://www.w3.org/2001/sw/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.



**SemText - Semantic Hypertext - Making Latent Semantics Blatant**  
<http://semtext.org/mambo/index.php>

**SIG SEMIS Semantic Web and Information Systems**  
<http://www.sigsemis.org/>

**SIMAC - Foafing the Music - Semantic Interaction with Music Audio Contents**  
<http://foafing-the-music.iaa.upf.edu/>

**SIMILE Project - Semantic Interoperability of Metadata and Information in unLike Environments**  
<http://simile.mit.edu/>

**SOAPAgent - An Open SOAP Directory**  
<http://soapagent.com/>

**SourceForge.net: Project Info - OWL API**  
<http://sourceforge.net/projects/owlapi>

**Swoogle - Semantic Bot**  
<http://swoogle.umbc.edu/>

**SWRL: A Semantic Web Rule Language Combining OWL and RuleML**  
<http://www.daml.org/2003/11/swrl/>

**Technology Review: Sir Tim Berners-Lee - The Semantic Web**  
<http://www.technologyreview.com/articles/04/10/frauenfelder1004.asp>

**The Cover Pages**  
<http://xml.coverpages.org/>

**The Memetic Web**  
<http://www.memeticweb.org/>

**The ontoprise® GmbH**  
<http://www.ontoprise.de/>

**The RDF Query Language (RQL)**  
<http://139.91.183.30:9090/RDF/RQL/>

**The Semantic Grid**  
<http://www.semanticgrid.org/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**The Semantic Social Network by Stephen Downes**

<http://www.downes.ca/cgi-bin/website/view.cgi?dbs=Article&key=1076791198>

**The Semantic Web: An Introduction**

<http://infomesh.net/2001/swintro/>

**The Semantic Web By Tim Berners-Lee, James Hendler and Ora Lassila**

<http://snipurl.com/297g>

**The Semantic Web In Breadth**

<http://logicerror.com/semanticWeb-long>

**The Semantic Indexing Project - Creating Tools To Identify the Latent Knowledge Found in Text**

<http://www.knowledgesearch.org/>

**The Semantic Web Is Your Friend**

<http://www.freepint.com/issues/270504.htm#feature>

**Transforming and Enriching Documents for the Semantic Web by Dietmar Roesner, Manuela Kunze, Sylke Kroetzsch**

<http://arxiv.org/abs/cs.AI/0501096>

**UDDI - Universal Description, Discovery, and Integration**

<http://www.uddi.org/>

**Web Semantics: Science, Services and Agents on the World Wide Web**

<http://www.sciencedirect.com/science/journal/15708268>

**Web Service Modeling Ontology**

<http://www.wsmo.org/>

**Wilbur Toolkit for Semantic Web Programming**

<http://wilbur-rdf.sourceforge.net/>

**WonderWeb**

<http://wonderweb.man.ac.uk/owl/>

**World Wide Web Reference**

<http://www.WWWReference.info/>

**XML.com: Semantic Web**

[http://www.xml.com/pub/rg/Semantic\\_Web](http://www.xml.com/pub/rg/Semantic_Web)



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**XML.org**

<http://www.xml.org/>

**Yahoo Groups - SemanticWeb**

<http://groups.yahoo.com/group/semanticweb/>

## **BOT RESEARCH RESOURCES AND SITES:**

**1st Spot**

[http://1st-spot.net/topic\\_agents.html](http://1st-spot.net/topic_agents.html)

**Agent-Based Software Development**

<http://www.ecs.soton.ac.uk/~mml/absd/index.html>

**Agent Construction Tools**

<http://www.agentbuilder.com/>

**AgentLand**

<http://www.agentland.com/>

**AgentLink**

<http://www.AgentLink.org/>

**Agent Model Yields Leadership**

<http://snipurl.com/99mh>

**Agent Portal AI**

<http://www.agent.ai/>

**Agents**

<http://www.aaai.org/AITopics/html/agents.html>

**Agents Portal**

<http://aose.ift.ulaval.ca/>

**Alarm Growing Over Bot Software by Robert Lemos**

[http://news.com.com/2100-7349\\_3-5202236.html?tag=nefd.led](http://news.com.com/2100-7349_3-5202236.html?tag=nefd.led)

**ALICEBot**

<http://www.alicebot.org/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Android World**

<http://www.androidworld.com/index.htm>

**Applied Soft Computing**

<http://www.sciencedirect.com/science/journal/15684946>

**B.4.1 Search Robots - The Robots.txt File**

<http://www.w3.org/TR/REC-html40/appendix/notes.html#h-B.4.1>

**Bot A Blog**

<http://www.BotABlog.com/>

**Botizen**

<http://www.botizen.com/>

**BotMakers**

<http://www.botmakers.com/>

**Bots, Blogs and News Aggregators**

<http://www.BotsBlogs.com>

**BotSpot®**

<http://www.botspot.com/>

**Build a Web Spider on Linux - A Simple Spider and Scraper Collects Internet Content**

<http://snipurl.com/128e6>

**Cetus Links - Mobile Agents**

[http://www.cetus-links.org/oo\\_mobile\\_agents.html](http://www.cetus-links.org/oo_mobile_agents.html)

**ChatterBots**

<http://www.ChatterBots.info/>

**Data Mining Resources**

<http://www.DataMiningResources.info/>

**Deep Web Research**

<http://www.deepwebresearch.info/>

**Design of a Parallel and Distributed Web Search Engine by Salvatore Orlando, Raffaele Perego, and Fabrizio Silvestri**

<http://arxiv.org/abs/cs.IR/0407053>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Dictionary of Algorithms and Data Structures**

<http://www.nist.gov/dads/>

**Eliza - The Original ChatterBot**

<http://www-ai.ijs.si/eliza/eliza.html>

**FAME (Facilitating Agents in Multiculture Exchange)Project**

<http://isl.ira.uka.de/fame/>

**Fantomas Spider Spy™ The BotBase**

<http://fantomaster.com/fasvsspy01.html>

**Foundation for Intelligent Physical Agents**

<http://www.fipa.org/>

**FyberSearch**

<http://www.fybersearch.com/>

**GeneSys Middleware**

<http://sourceforge.net/projects/genesys-mw/>

**Google Guide**

<http://www.googleguide.com/>

**Indexing Robot Crawler Checklist**

<http://www.searchtools.com/robots/robot-checklist.html>

**Institute for Human and Machine Cognition (IHMC)**

<http://www.ihmc.us/>

**Intellexer - Custom Built Search Engines, Knowledge Management Tools, Natural Language Processing**

<http://www.intellexer.com/>

**Internet Agents - CWS Apps**

<http://cws.internet.com/32agents.html>

**Internet Mathematics**

<http://www.InternetMathematics.org/>

**Journal of Mathematical Modelling and Algorithms**

<http://www.springerlink.com/link.asp?id=108992>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**KiwiLogic**

<http://www.kiwilogic.com/>

**Knowledge Discovery**

<http://www.knowledgediscovery.info/>

**Koders - Source Code Search Engine**

<http://koders.com/>

**LifeFX**

<http://www.lifefxtechnologies.com/>

**LAIR - Research Projects of the Laboratory of Applied Informatics Research**

<http://lair.indiana.edu/research/>

**List of User-Agents (Spiders, Robots, Crawler, Browser)**

<http://www.psychedelix.com/agents/index.shtml>

**Minimal-Intelligence Agents for Bargaining Behaviors in Market-Based Environments by Dave Cliff and Janet Bruten**

<http://www.hpl.hp.com/techreports/97/HPL-97-91.html>

**MIT Media Lab: Software Agents**

<http://agents.media.mit.edu/index.html>

**Modelling and Mining of Network Information Systems**

<http://www.mathstat.dal.ca/~mominis/index.html>

**MultiAgent**

<http://www.MultiAgent.com/>

**MySpiders**

<http://myspiders.informatics.indiana.edu/>

**Open Source Web Information Retrieval (OSWIR05)**

<http://www.emse.fr/OSWIR05/>

**Oxyus Search Engine**

<http://sourceforge.net/projects/oxyus/>

**PhpDig.net - Web Spider and Search Engine**

<http://www.phpdig.net/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Robots, Spiders and Other User Agents: A Resource for WebMasters**

<http://jose-luis.pellicer.org/ua/>

**Robots.Txt Checker - Validator for Robots.txt Files**

<http://tool.motoricerca.info/robots-checker.phtml>

**RobotsTxt.org**

<http://www.robotstxt.org/>

**Search Engine Robots**

<http://www.jafsoft.com/searchengines/webbots.html>

**Search Engine Watch News**

<http://www.searchenginewatch.com/>

**Search Tools - Information Guides and News**

<http://www.searchtools.com/>

**Semantic Indexing**

[http://www.nitle.org/semantic\\_search.php](http://www.nitle.org/semantic_search.php)

**Semantic Web**

<http://www.semanticweb.org/>

**ShoppingBots**

<http://www.ShoppingBots.info/>

**SiteMaps.org**

<http://www.SiteMaps.org/>

**Smarter Bots**

<http://www.SmarterBots.com/>

**SocSciBot3**

<http://socscibot.wlv.ac.uk/>

**Spider Hunter**

<http://www.spiderhunter.com/>

**SpiderMakers**

<http://www.spidermakers.com/index.shtml>

**Spidering Hacks**

<http://www.oreilly.com/catalog/spiderhks/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Structure and Interpretation of Computer Programs - Video Lectures by Hal Abelson and Gerald Jay Sussman**

<http://www.swiss.ai.mit.edu/classes/6.001/abelson-sussman-lectures/>

**Supybot, A Superb Python IRC Bot**

[http://freshmeat.net/projects/supybot/?branch\\_id=31808&release\\_id=181322](http://freshmeat.net/projects/supybot/?branch_id=31808&release_id=181322)

**Swoogle - Semantic Bot**

<http://swoogle.umbc.edu/>

**The CGI Resource Index: Programs and Scripts: Perl: Searching**

[http://cgi.resourceindex.com/Programs\\_and\\_Scripts/Perl/Searching/](http://cgi.resourceindex.com/Programs_and_Scripts/Perl/Searching/)

**The Intelligent Software Agents Lab**

<http://www-2.cs.cmu.edu/~softagents/>

**The Search Engine Project (TSEP)**

<http://freshmeat.net/projects/tsep/>

**The Simon Lavern Page**

<http://www.simonlaven.com/>

**The Web Robots Pages**

<http://www.robotstxt.org/wc/robots.html>

**UMBC AgentWeb**

<http://agents.umbc.edu/>

**UMBC eBiquity**

<http://ebiquity.umbc.edu/>

**Webbot - the W3C libwww Robot**

<http://www.w3.org/Robot/>

**Web Data Extractors - White Paper Link Compilation**

<http://zillman.blogspot.com/2004/09/web-data-extractors.html>

**Web Intelligence Consortium**

<http://wi-consortium.org/>

**Web IR & IE**

<http://www.webir.org/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.



## **Worm Radar**

<http://wormradar.com/index.html>

## **Subject Tracer™ Information Blogs:**

Subject Tracer™ Information Blogs created and developed by the Virtual Private Library™ combine the best of the latest tools on the Internet. Using bots, blogs and news aggregators the Subject Tracer™ Information blogs generate RSS feeds with the latest resources to create a current information resource flow through niched subject tracers. I am proud to be the creator of the Internet's first Subject Tracer™ Information Blogs:

### **Virtual Private Library**

<http://www.VirtualPrivateLibrary.com/>

### **Accessibility Resources**

<http://www.AccessibilityResources.info/>

### **Agriculture Resources**

<http://www.AgricultureResources.info/>

### **Artificial Intelligence Resources**

<http://www.AIResources.info/>

### **Astronomy Resources**

<http://www.AstronomyResources.info/>

### **Auction Resources**

<http://www.AuctionResources.info/>

### **Biological Informatics**

<http://www.biologicalinformatics.info/>

### **Biotechnology Resources**

<http://www.BiotechnologyResources.info/>

### **Bot Research**

<http://www.botresearch.info/>

### **Business Intelligence Resources**

<http://www.biresources.info/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**ChatterBots**

<http://www.ChatterBots.info/>

**Data Mining Resources**

<http://www.DataMiningResources.info/>

**Deep Web Research**

<http://www.deepwebresearch.info/>

**Directory Resources**

<http://www.DirectoryResources.info/>

**eCommerce Resources**

<http://www.eCommerceResources.info/>

**Elder Resources**

<http://www.ElderResources.info/>

**Employment Resources**

<http://www.EmploymentResources.info/>

**Entrepreneurial Resources**

<http://www.EntrepreneurialResources.info/>

**Financial Sources**

<http://www.FinancialSources.info/>

**Finding People**

<http://www.FindingPeople.info/>

**Games Resources**

<http://www.GamesResources.info/>

**Genealogy Resources**

<http://www.GenealogyResources.info/>

**Grant Resources**

<http://www.GrantResources.info/>

**Grid Resources**

<http://www.GridResources.info/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**Healthcare Resources**

<http://www.healthcareresources.info/>

**Information Futures Markets**

<http://www.InformationFuturesMarkets.com/>

**Information Quality Resources**

<http://www.InformationQualityResources.info/>

**Internet Alerts**

<http://www.InternetAlerts.info/>

**Internet Demographics**

<http://www.internetdemographics.info/>

**Internet Experts**

<http://www.internetexperts.info/>

**Internet Hoaxes**

<http://www.internethoaxes.info/>

**Journalism Resources**

<http://www.JournalismResources.info/>

**Knowledge Discovery**

<http://www.knowledgediscovery.info/>

**Military Resources**

<http://www.MilitaryResources.info/>

**Outsourcing/Offshoring Information and Resources**

<http://www.OutsourcingOffshore.us/>

**Privacy Resources**

<http://www.PrivacyResources.info/>

**Reference Resources**

<http://www.ReferenceResources.info/>

**Research Resources**

<http://www.researchresources.info/>



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>

[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.

**RestStress™**

<http://www.RestStress.com/>

**Script Resources**

<http://www.ScriptResources.info/>

**ShoppingBots**

<http://www.ShoppingBots.info/>

**Social Informatics**

<http://www.SocialInformatics.net/>

**Statistics Resources**

<http://www.statisticsresources.info/>

**Student Research**

<http://www.studentresearch.info/>

**Theology Resources**

<http://www.TheologyResources.info/>

**Tutorial Resources**

<http://www.TutorialResources.info/>

**World Wide Web Reference**

<http://www.WWWReference.info/>

Deep Web Research 2007 is a very exciting place to search and to do research. New tools are constantly being created, and more databases and unique files are constantly being added. This all adds up to a phenomenal growth area of the world wide web that deserves your constant attention through search and current awareness that keeps you alert for the latest happenings and sources available on the Internet! This article is constantly updated as its source is the Deep Web Research Subject Tracer™ Information Blog (<http://www.DeepWebResearch.info/>)

Marcus P. Zillman, M.S., A.M.H.A., is Executive Director of the Virtual Private Library and Founder/Creator of BotSpot®. He is the author of nine different Internet MiniGuides 2007 (<http://www.InternetMiniGuides.com/>), Internet Sources Manual (<http://www.InternetSources.info/>) and eCurrent Awareness Resources 2007 Report (<http://www.eCurrentAwareness.com>). His Subject Tracer™ Information Blogs (47 and constantly growing) are freely available from the Virtual Private Library (<http://www.VirtualPrivateLibrary.com>), which include the latest resources on Deep Web



Research (<http://www.DeepWebResearch.info>) and Bot Research (<http://www.BotResearch.info>). His current white papers on searching and researching the Internet are located at WhitePapers.us (<http://www.WhitePapers.us>). His personal blog with over 6000 postings is dedicated to knowledge discovery, knowledge harvesting, information retrieval and Internet current awareness and is available at Zillman.us (<http://www.Zillman.us>). His monthly free newsletter is titled AwarenessWatch™ (<http://www.AwarenessWatch.com>) and his monthly Internet Zillman Column has been archived since 1996 ( <http://www.ZillmanColumns.com>).



Awareness Watch V5N1 January 2007 Newsletter

<http://www.AwarenessWatch.com/>  
[zillman@VirtualPrivateLibrary.com](mailto:zillman@VirtualPrivateLibrary.com)

© 2007 Marcus P. Zillman, M.S., A.M.H.A.