Data Mining and Web Data Extractors 2012

By

Marcus P. Zillman, M.S., A.M.H.A.
Executive Director – Virtual Private Library
zillman@virtualprivatelibrary.com

This February 2012 column covers **Data Mining and Web Data Extractors 2012** and is a comprehensive listing of data mining and web data extractors resources and resources on the Internet that may be accessed freely. The below list of sources is taken from my Subject Tracer™ Information Blog titled Data Mining Resources and is constantly updated with Subject Tracer™ bots at the following URL:

http://www.DataMiningResources.info/

The Web Data Extractors are available from the white paper site titled White Papers by Marcus P. Zillman at the URL of http://WhitePapers.us/. These sources and resources will help you to discover the many pathways available through the Internet to find the latest data mining and web data extraction resources that are being used in today's highly competitive environment for the New and Existing Economy.

Data Mining and Web Data Extractors 2012

Data Mining:

80legs - Powerful and Economical Service Platform for Crawling and Processing Web Content
http://www.80legs.com/

ACM SIGKDD: Current Explorations Issue
http://www.acm.org/sigs/sigkdd/explorations/issue.php?issue=current

Anthracite Web Mining Desktop Toolkit for MacOS X
http://www.metafy.com

Apache Pig – Platform for Analyzing Large Datasets
http://pig.apache.org/

ARTstor - Digital Image Library for Education and Scholarship
http://www.artstor.org/

AZMY Thinkware -- Data Analysis and Mining Software Tools
http://www.azmy.com/

Benchmarking- Data Mining Benchmarking Association
http://www.dmbenchmarking.com/

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/

BI-DW - Business Intelligence and Data Warehousing Directory
http://www.bi-dw.info/

Biomedical LIterature (and text)Mining Publications (BLIMP)
http://blimp.cs.queensu.ca/

Bixo - Open Source Web Mining Toolkit
http://bixo.101tec.com

Bixolabs - Elastic Web Mining Platform
http://bixolabs.com

BLIASoft Knowledge Discovery
http://www.bliasoft.com/Eindex.html

Bologna Data Mining Centre
http://open.cineca.it/datamining/index_ing.htm

Bot Research
http://www.BotResearch.info/

Business Intelligence Books
http://www.BusinessIntelligenceBooks.com/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Business Intelligence Data Mining
http://businessintelligence.ittoolbox.com/nav/t.asp?t=312&p=312&h1=312

Business Intelligence Text Mining
http://businessintelligence.ittoolbox.com/nav/t.asp?t=316&p=316&h1=316

Business Intelligence Resources
http://www.BIResources.info/

Business Intelligence Web Mining
http://businessintelligence.ittoolbox.com/nav/t.asp?t=322&p=322&h1=322

CCSU - Data Mining
http://www.ccsu.edu/datamining/

Center for Automated Learning and Discovery
http://www.cs.cmu.edu/~cald/

ChartSearch - Intelligent Data Search
http://www.chartsearch.net/

Chronicling America - Library of Congress - National Digital Newspaper Program
http://www.loc.gov/chroniclingamerica/

Clementine Data Mining Workbench
http://www.spss.com/clementine

CRoss Industry Standard Process for Data Mining (CRISP-DM)
http://www.crisp-dm.org/

Current Awareness Discovery Tools on the Internet
http://zillman.blogspot.com/2004/09/current-awareness-discovery-tools-on.html

Cypher - Plain Language Acess to the Semantic Web
http://www.monrai.com/products/cypher

D2K - Data to Knowledge
http://alg.ncsa.uiuc.edu/do/tools/d2k

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Data Engineering Bulletin
http://tab.computer.org/tcde/bull_about.html

DataFerrett - Data Mining Tool
http://dataferrett.census.gov/

Data Fountains: Open Source Internet Resource Discovery and Metadata/Full-Text Generation Service
http://infomine.ucr.edu/Data_Fountains/

Data Mining
http://zillman.blogspot.com/2005/05/data-mining.html

Data Mining and Analytic Technologies
http://www.thearling.com/

Data Mining and KDD Papers
http://www.andypryke.com/university/papers.html

Data Mining and Knowledge Discovery Journal
http://snipurl.com/61lnx

Data Mining - Federal Efforts Cover a Wide Range of Uses Report
http://www.gao.gov/new.items/d04548.pdf

DataMiningGrid Consortium
http://www.datamininggrid.org/

Data Mining Group (DMG)
http://www.dmg.org/

Data Mining, Predictive Modeling, Business Analytics: Training, Consulting & Solutions
http://www.the-modeling-agency.com/

Data Mining Resources
http://www.cs.purdue.edu/homes/ayg/CS590D/resources.html

Data Mining Resources
http://datamining.togaware.com/

zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Data Mining Resources at CCSU
http://www.ccsu.edu/datamining/resources.html

Data Mining Sifts the Gems From Digital Ore By Wilson P. Dizard III
http://www.gcn.com/23_29/news/27420-1.html

Data Mining: Technology and Policy The DHS Privacy Office
http://www.dhs.gov/xlibrary/assets/privacy/privacy_rpt_datamining_200812.pdf

Data Mining: Text Mining, Visualization and Social Media
http://datamining.typepad.com/data_mining/

Data Mining Using SAS Enterprise Miner by Randall Matignon
http://www.sasenterpriseminer.com/

Data-Mining Virtual Machines for Resource Optimization
http://groups.csail.mit.edu/EVO-DesignOpt/evo.php?n=Site.SysML

Data Mining, Web Scraping, Web Mining, Data Extraction and Screen Scraping Technology Links
http://www.connotate.com/

Data Mining, Web Mining, and Business Intelligence Solutions from Salford Systems
http://www.salford-systems.com/

Data Mining, Web Mining and Knowledge Discovery Resources
http://www.eruditionhome.com/datamining/

Data-PASS
http://www.icpsr.umich.edu/DATAPASS/

Data Science Toolkit
http://www.datasciencetoolkit.org/

Data Shaping Data Mining Resources
http://www.datashaping.com/data_mining.shtml

Data Sources
http://www.andypryke.com/university/datasources.html

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

DataSpace - Web for Data
http://www.dataspaceweb.org/

DbVisualizer - The Universal Database Tool
http://www.dbvis.com/products/dbvis/

Deep Web Research
http://www.DeepWebResearch.info/

Dig Deep: Data mining Resources for Consultants by Beth Blakely
http://snipurl.com/6ohs

DigiCULT Resources - Resource Discovery & Information Retrieval
http://www.digicult.info/pages/resources.php?t=21

digitalAGORA
http://aut.edu/agora/

Digital Library for Earth System Education (DLESE)
http://www.dlese.org/

Directory of Data Warehouse, Data Mining, and Decision Support Resources
http://www.infogoal.com/dmc/dmcdwh.htm

DiscoverText - Unlock the Power of Text
http://discovertext.com/

DM Review
http://www.dmreview.com/

eBiquity Research Group Blogger
http://ebiquity.umbc.edu/v2.1/blogger/

Early Canadiana Online
http://www.canadiana.ca/

Elastic Web Mining Talk
http://www.slideshare.net/kkrugler/elastic-web-mining-2407818

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Emping - Data Mining Tool
http://j-van-thiel.speedlinq.nl/emp/empug.html

English Broadside Ballad Archive (EBBA)
http://ebba.english.ucsb.edu/

Enterprise Semantic Intelligence™ Knowledge Suite
http://www.transinsight.com/products

Exclusive Ore, Inc.
http://www.xore.com/

FACTA+ - Finding Associated Concepts with Text Analysis
http://refine1-nactem.mc.man.ac.uk/facta/

Four-T-Nine-R(sm): Data Mining in Web and non-Web Bibliographic Databases
http://www.public.iastate.edu/~CYBERSTACKS/4T9R.htm

Genalytics - Advanced Analytics for Marketing and Risk Management
http://www.genalytics.com/

GeneMiner
http://www.biomedcentral.com/1471-2105/8/S8/P3

German Data Mining Portal
http://mitglied.lycos.de/hpn/DataMining.html

GMDH - Group Method of Data Handling
http://come.to/GMDH

Google Refine 2.0 – Power Tool for Data Wranglers
http://code.google.com/p/google-refine/

Graf-FX - Visual Database Data Mining Software
http://www.vb123.com/graf/

Great War Primary Documents Archive
http://www.gwpda.org/

zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Harvard Time Series Center (TSC)
http://timemachine.iic.harvard.edu/search/

Howard D. Wactlar Home Page
http://www-2.cs.cmu.edu/~hdw/

Imagination Engines
http://www.Imagination-Engines.com/

InfoBionics - Flexible Data Mining Applications
http://www.infobionics.com/

Infochimps.org
http://infochimps.org/

Information Retrieval (IR) and Information Extraction (IE) on the Web Using Hypertext Meta-Data and Structure
http://www.webir.org/

Information Retrieval Intelligence
http://www.miislita.com/

InfoVis CyberInfrastructure
http://iv.slis.indiana.edu/index.html

Insight Consulting
http://www.deej.com/insight/

Integrating Data Mining, Databases and Information Retrieval (IDDI-05)
http://iddi05.unibg.it/

International Journal of Business Intelligence and Data Mining (IJBIDM)
http://www.inderscience.com/ijbidm

International Journal of Data Mining and Bioinformatics (IJDMB)
http://www.inderscience.com/ijdmb

International Journal of Data Warehousing and Mining (IJDWM)
http://www.igi-global.com/journals/details.asp?id=4291

zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Internet Archive
http://www.archive.org/

Inter-university Consortium for Political and Social Research (ICPSR)
http://www.icpsr.umich.edu/

Journal of Data Mining and Knowledge Discovery
http://www.bioinfo.in/contents.php?id=42&page=aim

Junar - Discovering Data
http://www.junar.com/

Kaggle - Data Mining, Forecasting and BioInformatics Competitions
http://kaggle.com/

KDD-2008
http://www.kdd2008.com/

KDD-2009
http://www.acm.org/sigs/sigkdd/kdd2009/

KDD-2010
http://www.sigkdd.org/kdd2010/

KDD-2011
http://www.sigkdd.org/kdd2011/

KDD-2012
http://www.sigkdd.org/kdd2012/

KDnuggets: Data Mining, Web Mining, and Knowledge Discovery Guide
http://www.kdnuggets.com/

KNIME – Konstanz Information Miner Open Source Software
http://www.knime.org/

Knowledge Discovery Resources
http://www.KnowledgeDiscovery.info/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Knowledge Discovery Resources 2012 Annotated White Paper Link Compilation by Marcus P. Zillman, M.S., A.M.H.A.
http://www.KDResources.info/

KnowleSys - Web Data Extraction
http://www.knowlesys.com

LingPipe - Information Extraction and Data Mining Tools
http://alias-i.com/lingpipe/

LLRX - A Review of TRACFed: Lawyers Strike Gold Mining Government Data
http://www.llrx.com/features/tracfed.htm

LLRX - Deep Web Research 2012
http://zillman.blogspot.com/2011/02/llrx-february-2011-issue-deep-web.html
http://DeepWeb.us/

Marcus P. Zillman Home Page
http://www.zillman.us/

Marriott Library at the University of Utah Digital Collections
http://www.lib.utah.edu/portal/site/marriottlibrary/

Marti Hearst Home Page
http://www.sims.berkeley.edu/~hearst/

Media Patterns - Detecting Patterns in the Global Media Content
http://mediapatterns.enm.bris.ac.uk/

MedScan - Automated Scientific Text Mining Tool
http://www.ariadnegenomics.com/products/medscan.html

Megaputer - Data Mining, Text Mining and Web Mining Software
http://www.megaputer.com/

Metafy Anthracite Web Mining Desktop For MACOS X
http://www.metafy.com

Microsoft® Data Mining Project - Efficient Data Exploration and Modeling
http://research.microsoft.com/dmx/DataMining/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Mi Li Wo Data Mining Community
http://www.geocities.com/misforto/

MineKnowledge – Revealing Your Data's Secrets
http://mineknowledge.com/

MonetDB Query Processing at Light Speed
http://monetdb.cwi.nl/

Mozenda - Comprehensive Web Data Gathering
http://www.mozenda.com/

National Archives, London
http://nationalarchives.gov.uk/

National Centre for Text Mining (NaCTeM)
http://www.nactem.ac.uk/

National Science Digital Library (NSDL)
http://www.nsdl.org/

National Technical Information Service (NTIS)
http://www.ntis.gov/

Nebraska Digital Newspaper Project
http://cdrh.unl.edu/nebnewspapers/

Nesstar
http://www.nesstar.com/

NetOwl - Discovery Software from SRA International
http://www.netowl.com/index.html

NewsTin - Multilingual News Search
http://www.newstin.com/

New York Public Library
http://www.nypl.org/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Nuix - eDiscovery and Electronic Investigation Software
http://www.nuix.com/

Oceanstore Project
http://oceanstore.cs.berkeley.edu/

Office of the Director of National Intelligence Data Mining Report (Unclassified)
http://www.dni.gov/reports/data_mining_report_feb08.pdf

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

Open Access Now - Data Mining Open Access Research
http://snipurl.com/6o2r

Open Directory Project - Data Mining
http://www.dmoz.org/Computers/Software/Databases/Data_Mining/

Opening History (OH) - U.S. History Resources from Libraries, Museums, and Archives
http://imlsdcc.grainger.uiuc.edu/history/

Open/Public Data Sources
http://bixolabs.com/datasets/public-datasets/

Open Source Data Mining Tools
http://bixolabs.com/oss/open-source-data-mining-tools/

Open Source Data Warehousing
http://www.infobright.com/index.php

Orange - Data Mining Software
http://www.ailab.si/orange

PC AI Magazine
http://www.pcai.com/pcai

Pentaho BI Project - Open Source Business Intelligence
http://www.pentaho.org/

PEPITe S.A. - Unlock Your Knowledge
http://www.pepite.be/

Prediction Markets
http://www.PredictionMarkets.com/

Predictive Model Markup Language (PMML)- SourceForge.net: Project Info
http://sourceforge.net/projects/pmml

Predictive Model Markup Language (PMML)
http://www.oasis-open.org/cover/pmml.html

Professional Web Data Extraction and Web Data Mining Resources and Solutions
http://www.web-extraction.com/

PubGene™ Database and Tools
http://www.pubgene.org/

Pudget - Science at Speed
http://corporate.pubget.com/

QL2 Software - Unstructured Data Management and Web Mining Software
http://www.ql2.com/

Raghu Ramakrishnan Home Page
http://www.cs.wisc.edu/~raghu/

RapidMiner - Open Source Data Mining Tool
http://rapid-i.com/content/blogcategory/10/69/

reSearcher
http://researcher.sfu.ca/

Rexer Analytics - Data Mining and CRM Analytics
http://www.rexeranalytics.com/

Ron Kohavi Home Page
http://robotics.stanford.edu/~ronnyk/

SAS - Data and Text Mining
http://snipurl.com/6o33

Scholarly Database at the Cyberinfrastructure for Network Science Center, Indiana University
http://sdb.slis.indiana.edu/

Sciengy RPF!™
http://sciengy.com/home

ScrapeGoat
http://www.scrapegoat.com/

Screen-Scraper - Extracts Information From Web Sites
http://www.screen-scraper.com/

Searching the Internet
http://www.SearchingTheInternet.info/

SIGKDD - ACM Special Interest Group - Knowledge Discovery in Data and Data Mining
http://www.acm.org/sigs/sigkdd/

Smithsonian/NASA Astrophysics Data System (ADS)
http://ads.harvard.edu/

Software Suites for Data Mining, Analytics, and Knowledge Discovery
http://www.kdnuggets.com/software/suites.html

SpagoBI - Unified Open Source Platform for Business Intelligence
http://www.spagoworld.org/ecm/faces/public/guest/home/solutions/spagobi

SPIDER: Scalable, Parallel and Interactive Data Mining and Exploration at Rensselaer
http://www.cs.rpi.edu/~zaki/datamining.html

Special Interest Group - Knowledge Discovery in Data and Data Mining - SIGKDD Explorations Newsletter
http://www.acm.org/sigs/sigkdd/explorations/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

SPSS, Data Mining, Statistical Analysis Software, Predictive Analysis, Predictive Analytics, Decision Support Systems
http://www.netgen.com/

SQL Server Data Mining
http://www.sqlserverdatamining.com/

Statistical Analysis and Data Mining
http://www3.interscience.wiley.com/journal/112701062/home

Statistical Data Mining Tutorials - Tutorial Slides by Andrew Moore
http://www-2.cs.cmu.edu/~awm/tutorials/index.html

Statoo Consulting Data Mining Resources Links
http://www.statoo.com/en/resources/anthill/Datamining/

Survey of DHS Data Mining Activities - Office of Information Technology
http://snipurl.com/wgff

T2K - Text to Knowledge
http://alg.ncsa.uiuc.edu/do/tools/t2k

Talend Open Data Solutions
http://www.talend.com/

Text Data Mining
http://www.sims.berkeley.edu/~hearst/talks/dm-talk/

Text Mining for Scholarly Communications and Repositories
http://www.nactem.ac.uk/tm-ukoln.php

Text Mining, Web Mining, Information Retrieval and Extraction from the WWW References
http://filebox.vt.edu/users/wfan/text_mining.html

The Archaeology Data Service (ADS)
http://ads.ahds.ac.uk/

The Centre for Contemporary Canadian Art - Canadian Art Database Project
http://www.ccca.ca/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

The Data Mine
http://www.the-data-mine.com/

The Data Warehouse
http://www.datawarehouse.com/

The History Data Service (HDS)
http://hds.essex.ac.uk/

The National Center for Data Mining
http://www.ncdm.uic.edu/

The National Centre for Text Mining: Aims and Objectives by Sophia Ananiadou, Julia Chruszcz, John Keane, John McNaught and Paul Watry
http://www.ariadne.ac.uk/issue42/ananiadou/

The New York Times Article Search API
http://developer.nytimes.com/

The Open Access Digital Library
http://grweb.coalliance.org/oadl/oadl.html

Togaware - Data Mining Resources
http://datamining.togaware.com/

Topic Detection and Tracking (TDT)
http://www.itl.nist.gov/iaui/894.01/tests/tdt/

T-Rex (Trainable Relation Extraction)
http://sourceforge.net/projects/t-rex/

Truthy - Analyzed and Visualize the Diffusion of Information on Twitter
http://truthy.indiana.edu/

twendz - Mining and Exploring Twitter Conversations and Sentiment
http://twendz.waggeneredstrom.com/

Unit Miner - Web Data Extraction Software
http://www.qualityunit.com/unitminer/web-extraction-tool.html

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

University of Florida Digital Collections (UFDC)
http://www.uflib.ufl.edu/ufdc

University of North Texas Digital Collections
http://digital.library.unt.edu/browse/?browseby=collection

Using the Internet As a Dynamic Resource Tool for Knowledge Discovery
http://zillman.blogspot.com/2009/08/using-internet-as-dynamic-resource-tool.html

Vendor-Neutral Public Courses on Data Mining Strategy, Methods & Practice
http://www.the-modeling-agency.com/training

VisitorVille - Web Site Intelligence
http://www.visitorville.com/

Visual Analytics
http://www.visualanalytics.com/

Web Curator Tool (WCT)- Management of Selective Web Harvesting Process
http://webcurator.sourceforge.net/

Web Data Extractors - White Paper Link Compilation
http://www.WebDataExtractors.com/

Web Data Mining
http://www.blossom.com/web_mining.html

WebFarming.com - Farming the Web for Systematic Business Intelligence
http://www.webfarming.com/

Web-Harvest - Open Source Web Data Extraction Tool
http://web-harvest.sourceforge.net/index.php

Web Harvesting by Russell Kay
http://snipurl.com/78sy

Web Mining
http://galeas.de/webmining.html

Web Mining - Business Intelligence
http://snipurl.com/6ogb

Weka 3 - Data Mining with Open Source Machine Learning Software in Java
http://www.cs.waikato.ac.nz/~ml/weka/index.html

White Papers by Marcus P. Zillman, M.S., A.M.H.A.
http://www.WhitePapers.us/

WizSoft - Data and Text Mining
http://www.wizsoft.com/

WordsCloseTogether.com - Meaningful Precision Search for Text Data
http://www.wordsclosetogether.com/

Yahoo Groups - Data Mining
http://groups.yahoo.com/group/datamining2/
http://groups.yahoo.com/group/datamining/

Web Data Extractors:

140kit - Free Open Source Twitter Analytics Platform
http://140kit.com/

80legs - Powerful and Economical Service Platform for Crawling and Processing Web Content
http://www.80legs.com/

Anomic HTTP Proxy
http://freshmeat.net/releases/173068/

Anthracite
http://freshmeat.net/projects/anthracite/

Automated RSS Scraper Scripts
http://www.djeaux.com/rss/

Automated Information Solutions
http://www.automated-info-solutions.com/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Automatic Information Extraction From Semi-Structured Web Pages By Pattern Discovery
http://portal.acm.org/citation.cfm?id=640423&dl=ACM&coll=portal

Automation Anywhere – Web Data Extraction Software
http://www.automationanywhere.com/solutions/webDataExt.htm

Beautiful Soup
http://freshmeat.net/projects/beautifulsoup/

Beautiful Soup - HTML/XML Parser for Quick Turnaround Screen Scraping and Web Data Extraction
http://www.crummy.com/software/BeautifulSoup/

BLIASoft Knowledge Discovery
http://www.bliasoft.com/Eindex.html

Bot Research
http://www.BotResearch.info/

BYU Data Extraction Research Group
http://www.deg.byu.edu/

Captiva Software: Digital Information Capture Software
http://www.captivasoftware.com/index.asp

ChartSearch Data Search Technology
http://www.ChartSearch.net/

Client-Side Deep Web Data Extraction
http://www.tic.udc.es/~mad/publications/ceceast2004.pdf

Connotate - Intelligent Agent Technology and Business Intelligence Tools
http://www.connotate.com/intelligent_software_agents.aspx

ContextMiner - Tools to Collect Data, Metadata and Contextual Information
http://www.contextminer.org/

cQuery - Content Query Engine
http://cquery.com/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Crowbar – Open Source Web Scraping Environment
http://simile.mit.edu/wiki/Crowbar

Dapper - Extract and Use Website Information for Mashups
http://www.dappit.com/index.php

Data Extraction Services
http://www.dataextractionservices.com/

Data Extractors and Pass-Through Systems – A Selected List
http://www.chass.utoronto.ca/datalib/misc/dli/extracts.htm

Data Extractors – Special Report
http://snipurl.com/8flw

Data Mining Resources
http://www.DataMiningResources.info/

DataWrangler - Data Cleaning and Transformation Tool
http://vis.stanford.edu/wrangler/

Deep Web Research
http://www.DeepWebResearch.info/

DiscoverText - Import, Sort, Distribute and Analyze Electronic Content from eMail, Document Repositories, and Social Media
http://discovertext.com/

Effective Web Data Extraction with Standard XML Technologies
http://www10.org/cdrom/papers/102/

eGrabber - Data Capture Tools
http://www.egrabber.com/

ExtractData Technologies - SearchExtract Software
http://www.extradata.com/

Extracting Knowledge
http://www.intelligentkm.com/feature/010507/feat1.jhtml?_requestid=185742

FeedsAPI - Extract Content from Web Pages Tool
http://www.feedsapi.com/

FerretSoft
http://www.FerretSoft.com/

Fetch Live Access
http://www.fetch.com/

Ficstar Software - Web Data Extraction
http://www.ficstar.com/index.html

File Information Tool Set (FITS)
http://code.google.com/p/fits/

Fresh WebSuction
http://www.freshwebmaster.com/

Happy Harvester – Advanced Web Data Extractor
http://www.happyharvester.com/

Imagination Engines
http://www.Imagination-Engines.com/

InfoExtractor - Extracts Relevant Information from Blogs, YouTube and Twitter
http://www.infoextractor.org/

Information Retrieval (IR) and Information Extraction (IE) on the Web
http://www.webir.org/

Information Retrieval Resources
http://www-csli.stanford.edu/~hinrich/information-retrieval.html

InfoSquire – Web Data Extraction
http://www.InfoSquire.com/

Introduction to Information Retrieval
http://informationretrieval.org/

iOpus® Internet Macros™
http://www.iopus.com/iim.htm

iRobotSoft – Visual Web Scraping and Web Automation
http://irobotsoft.com/

iWebMiner – Web Data Mining and Content Scraping
http://www.iwebminer.com/

Junar - Discovering Data
http://www.junar.com/

Kapow RoboSuite Platform - Solutions for Data Collection
http://www.kapowtech.com/

Kapow Web Collector
http://www.automated-info-solutions.com/

Knowledge Discovery Resources
http://www.KnowledgeDiscovery.info/

Knowlesys® - Web Data Extraction, Web Grabber and Screen Scraper
http://www.knowlesys.com/index.htm

LingPipe – Information Extraction and Data Mining Tools
http://alias-i.com/lingpipe/

Lingway
http://www.lingway.fr/en/

Mastering Data Extraction
http://www.dbmsmag.com/9606d05.html

Metadata Extraction Tool
http://meta-extractor.sourceforge.net/

mnot: xpath2rss - HTML->RSS scraper
http://www.mnot.net/xpath2rss/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Mozenda – Comprehensive Web Data Gathering
http://www.mozenda.com/

Newprosoft – Web Data Extraction Software
http://newprosoft.com/

NewsClipper.com - Snip and Ship Dynamic News Content to Your Web Pages
http://www.newsclipper.com/

Pervasive Data Management and Integration Products
http://www.pervasive.com/

Professional Web Data Extraction and Web Data Mining Solutions and Services
http://www.web-extraction.com/

QL2 Software - Unstructured Data Management and Web Mining Software
http://www.ql2.com/

REBOL Technologies
http://www.rebol.com/

ScrapeForge
http://freshmeat.net/projects/scrapeforge/

ScrapeGoat
http://www.ScrapeGoat.com/

ScrapePro – Universal Web Scraper Platform
http://scrapepro.com/

Scraper
http://freshmeat.net/projects/scraper/

ScraperWiki - Community of Programmers Sifting Information To Give You the Edge
https://scraperwiki.com/

Scrapy – Open Source Web Scraping Framework for Python
http://scrapy.org/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Screen-Scraper
http://freshmeat.net/projects/screenscraper/

Screen-Scraper – Extracts Information From Web Sites
http://www.Screen-Scraper.com/

Screenscraping the Senate by Paul Ford
http://www.xml.com/pub/a/2004/09/01/hack-congress.html

Screen Snarfs - Darkspell: Perl Code for Screen Scrapers
http://www.darkspell.com/gadgets/snarfs/

Search and Replace with TextPipe Pattern Matching
http://www.crystalsoftware.com.au/textpipe.html

Social Science Data Extractors
http://www.ssc.wisc.edu/cde/datalib/extract.htm

Text Mining and Web-Based Information Retrieval Reference
http://filebox.vt.edu/users/wfan/text_mining.html

Unit Miner - Web Data Extraction Software
http://www.qualityunit.com/unitminer/web-extraction-tool.html

URL Link Extractor
http://www.fuddyduddy.connectfree.co.uk/urlgen.htm

Visual Web Spider
http://www.newprosoft.com/

Visual Web Task
http://www.lencom.com/VisualWTSite.html

W3C Publishes Data Extraction Language (DEL) as W3C Note
http://xml.coverpages.org/ni2001-11-06-a.html

Web Data Extraction Deamon
http://webdatascraper.net/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Web Data Extraction Software
http://www.tethyssolutions.com/web-data-extraction.htm

Web Data Extractor
http://www.rafasoft.com/

Web Data Extractors
http://www.webextractors.com/

Web Data Extractor
http://www.rafasoft.com/

Web Data Mining
http://www.blossom.com/web_mining.html

Web Grabber
http://www.ficstar.com/web_grabber.html

Web-Harvest – Open Source Web Data Extraction Tool
http://web-harvest.sourceforge.net/index.php

Web Mining and Unstructured Data Management Solutions – QL2 Software
http://www.ql2.com/

WebQL
http://www.ig.com.au

WebScraper Plus +
http://www.velocityscape.com/

Website Extractor
http://www.hot-shareware.com/internet-tools/website-extractor/

Website Extractor – Offline Browser
http://www.internet-soft.com/extractor.htm

Website Scraping
http://www.websitescraping.com/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Web Spider, Link Extraction, And Other Extractor Products
http://www.pjltechnology.com/

WebSunDew – Advanced Web Scraping Tool
http://www.websundew.com/

Wikimedia Public Data Dumps
http://meta.wikimedia.org/wiki/Data_dumps

Words, Extended - Internet Text Information Rretrieval, Extraction and Display Bot
http://home.earthlink.net/~glenn_scheper/

XRay Web Scraping Tool
http://freshmeat.net/projects/xrayguibasedwebscrapingtool/

zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

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/

Agriculture Resources
http://www.AgricultureResources.info/

AnswerSpot
http://www.AnswerSpot.us/

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/

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://eCommerceResources.info/

Elder Resources
http://www.ElderResources.info/

Employment Resources
http://www.EmploymentResources.info/

Entrepreneurial Resources
http://www.EntrepreneurialResources.info/

Fact Checkers Directory
http://www.FactCheckers.us/

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/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Green Files
http://www.GreenFiles.info/

Grid, Distributed and Cloud Computing Resources
http://www.GridResources.info/

Healthcare Resources
http://www.HealthcareResources.info/

Information Futures Markets
http://www.InformationFutureMarkets.com/

Information Quality Resources
http://www.InformationQualityResources.info/

International Trade Resources
http://www.InternationalTradeResources.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/

Intrapreneurial Resources
http://www.IntrapreneurialResources.info/

Journalism Resources
http://www.JournalismResources.info/

Knowledge Discovery
http://www.KnowledgeDiscovery.info/

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Military Resources
http://www.MilitaryResources.info/

New Economy Analytics, Resources and Alerts
http://www.NewEconomyAnalytics.com/

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/

RestStress™
http://www.RestStress.com/

Script Resources
http://www.WcriptResources.info/

ShoppingBots
http://www.ShoppingBots.info/

Social Informatics
http://www.SocialInformatics.info/

Statistics Resources
http://www.StatisticsResources.info/

Student Research
http://www.StudentResearch.info/

Theology Resources
http://www.TheologyResources.info/

zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Tutorial Resources
http://www.TutorialResources.info/

World Wide Web Reference
http://www.WWWReference.info/

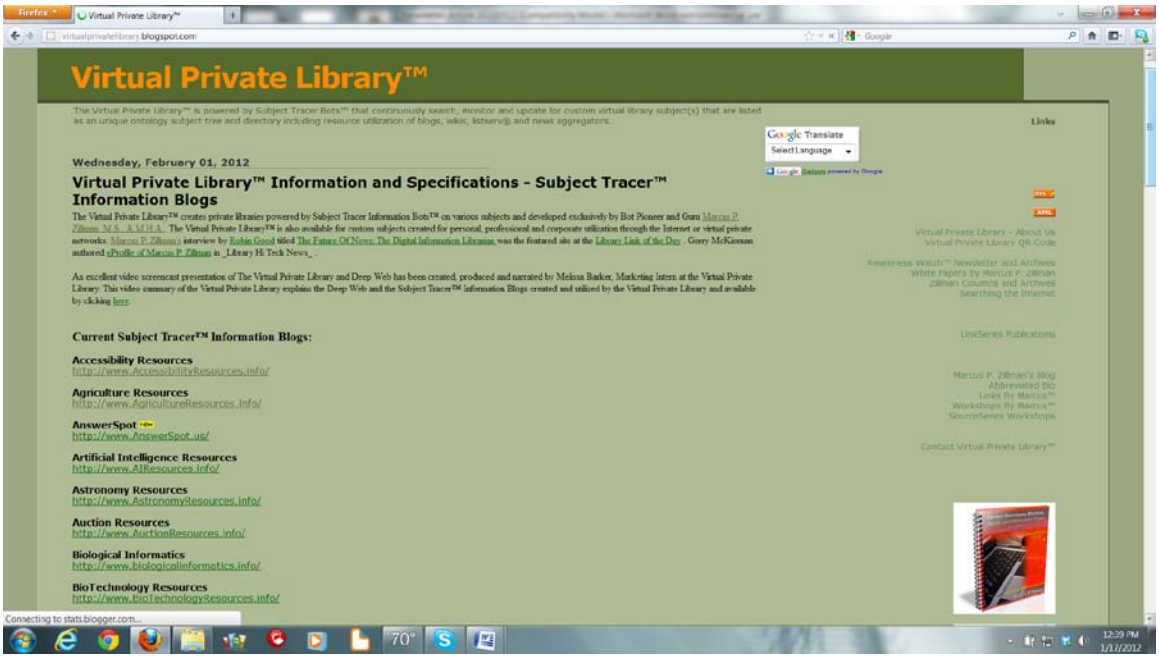

Figure 1: Virtual Private Library™

Author Information: Marcus P. Zillman, M.S., A.M.H.A. Executive Director of the Virtual Private Library is an international Internet expert, author, keynote speaker and corporate consultant in the area of information retrieval, knowledge discovery, knowledge harvesting, artificial intelligence and bots/intelligent agents. He has created numerous world wide web sites including 53 Subject Tracer™ Information Portals and Blogs; written a number of internet miniguides, white papers, manuals and books; hosted over 160 weekly Internet television shows, writes a weekly and monthly column on Current Awareness on the Internet; writes a monthly newsletter Awareness Watch and delivers keynote presentations throughout the international marketplace. He also actively delivers one and two day workshops for key industry sectors displaying how the Internet can be used as a tool to maintain current awareness and professional competencies.

Additional websites by Marcus P. Zillman, M.S., A.M.H.A.:

Marcus P. Zillman's Blog
http://www.zillman.us/

Marcus P. Zillman Abbreviated Bio
http://www.zillman.info/

White Papers by Marcus P. Zillman
http://www.WhitePapers.us/

Internet MiniGuides™
http://www.InternetMiniguide.com/

Awareness Watch™ Newsletter
http://www.AwarenessWatch.com/

Marcus P. Zillman's Columns
http://www.ZillmanColumns.com

LinkSeries Publications
http://www.LinkSeries.com/

Internet Sources™ Manual
http://www.InternetSources.info/

Links By Marcus™
http://www.LinksByMarcus.com/

Workshops By Marcus™
http://www.WorkshopsByMarcus.com/

SourceSeries Internet Research Workshops
http://www.SourceSeries.com/

Watch Marcus™
http://www.WatchMarcus.com/

listen to marcus™
http://www.ListenToMarcus.com

Research White Papers, Articles, Lectures and Speeches by Marcus P. Zillman, M.S., A.M.H.A.:

Academic and Scholar Search Engines and Sources
http://www.ScholarSearchEngines.com/

Bots, Blogs and News Aggregators
http://www.BotsBlogs.com/

Business Intelligence Online Resources
http://www.BIOnlineResources.info/

Cloud Computing Resources Primer
http://zillman.blogspot.com/2011/05/grid-distributed-and-cloud-computing.html

Current Awareness Discovery Tools on the Internet
http://zillman.blogspot.com/2009/08/current-awareness-discovery-tools-on.html

Deep Web Research 2012 Article - LLRX and Online White Paper
http://zillman.blogspot.com/2012/01/deep-web-research-2012.html
http://DeepWeb.us/

eReference Library Link Toolkit
http://www.eReferenceLibrary.com/

Finding Experts By Using the Internet
http://www.FindingExperts.info/

Finding People Resources and Sites
http://www.FindingPeople.info/

Healthcare Bots and Subject Directories
http://www.HealthcareBots.info/

Knowledge Discovery Resources 2012
http://www.KDResources.info/

Online Research Browsers
http://zillman.blogspot.com/2009/08/online-research-browsers.html

zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Online Research Tools
http://www.OnlineResearchTools.info/

Online Social Networking
http://zillman.blogspot.com/2009/08/online-social-networking.html

Searching the Internet
http://www.SearchingTheInternet.info/

Using the Internet As a Dynamic Resource Tool for Knowledge Discovery
http://zillman.blogspot.com/2009/08/using-internet-as-dynamic-resource-tool.html

Web Data Extractors
http://www.WebDataExtractors.com/

Web Guide for the New Economy
http://www.WebGuideNewEconomy.com/

White Papers By Marcus P. Zillman, M.S., A.M.H.A.
http://www.WhitePapers.us/

Internet Tutor by Marcus P. Zillman, M.S., A.M.H.A.
http://www.InternetTutor.info/
Visit this site to learn about the availability of Marcus P. Zillman to tutor you or your associate one on one in the privacy of your residence or office on the latest happenings of the Internet including Internet basics to advanced Internet searching using bots and creating your own personal blog

Internet Speaking by Marcus P. Zillman, M.S., A.M.H.A.
http://www.InternetSpeaker.net
Visit this site to learn about Marcus P. Zillman's speaking engagements for your organization meetings and events. View and listen to his previous presentations as well as his weekly television shows

Internet Consulting by Marcus P. Zillman, M.S., A.M.H.A.
http://InternetConsultant.BlogSpot.com/
Visit this site to obtain information about obtaining the consultation services of Marcus P. Zillman for your company including eCommerce audits, utilization of bots, blogs and news aggregators or the creation of your own personal virtual private library powered by Subject Tracer™ Information bots!

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462

Current Awareness Monitors, Alerts and Information Traps for 2010
http://www.ecurrentAwareness.com/
Marcus P. Zillman's latest report Current Awareness Monitors, Alerts and Information Traps for 2010 is now available for purchase online and for immediate download. This report is a comprehensive listing of the latest resources, sources and sites for current awareness on the Internet. This is a must read for anyone who must stay current in their profession and/or business activity as the list of URLs will keep you at the leading edge of your career.

Market Intelligence Resources 2010
http://www.MarketIntelligenceResources.com/
Marcus P. Zillman's just released professional Internet MiniGuide is titled Market Intelligence Resources 2010 and is now available for purchase online and immediate download. This 193 page digital miniguide represents a comprehensive listing of the latest resources, sources and sites to discover the latest Market Intelligence sources available on the Internet with many of them freely available! Designed specifically for today's entrepreneur, professional and/or investor.

Entrepreneurial Links 101
http://www.EntrepreneurialLinks.com/
Marcus P. Zillman's newly released 231 page eReference digital book for the up and coming entrepreneur. Entrepreneurial Links 101 gives an alphabetical listing of the very best Internet and World Wide Web sites covering Entrepreneur Resources, Business Intelligence Resources and an extremely comprehensive list of Online Research Tools. This is considered by many to be the entrepreneur's bible for finding relevant and competent online resources!

Internet Privacy and Security Resources
http://www.InternetPrivacySecurity.net/
Marcus P. Zillman's latest eReference digital publication is a selected comprehensive alphabetical listing of the latest resources and sites covering all aspects of privacy and security currently available over the Internet. From the board room to the family room, these resources and sites give you the information you need to maintain your privacy and security as you use the Internet in your business and personal life.

Research Resources Online Guide
http://www.ResearchResourcesOnline.net/
Marcus P. Zillman's latest LinkSeries Publication is a 340 page digital guide of a selected comprehensive alphabetical listing of the latest and greatest resources and sites covering all areas of research that is currently available over the Internet. The guide covers online

research resources and tools for the Newbie to research as well as the Seasoned researcher. Contents include: a) Research Resources, b) Research Tools, c) Student Research Resources Toolkit, d) Knowledge Discovery/Management and Data Mining Resources, e) Knowledge Discovery/Retrieval and the World Wide Web Resources, f) Business Intelligence Resources, g) Reference Resources, and h) Subject Tracer™ Information Blogs.

The Survivor's Manual for The New Economy.
http://www.NewEconomyManual.com/
Marcus P. Zillman's latest LinkSeries Publication is a 239 page digital read that gives excellent resources and annotated sources for the new economy analytics, alerts, ecommerce, financial sources, invisible and deep web resources, social and business networking sources along with new economy competitive and business intelligence resources and an extremely comprehensive listing of new economy online tools.

http://www.zillmancolumns.com/
zillman@VirtualPrivateLibrary.com
eVoice: (800) 858-1462