Data Mining Resources for Designers and Developers

Linda offers deals for designers and developers to mine their resources and to develop markets for products and services.

Keeping Tabs on Your Data

In the previous article, I covered reasons to mine and store data for analysis and for resources. In this article, I'll offer dozens of tools that can help you keep tabs on your own resources and clients. But, before I offer those tools, I want to hand over a list of blogs that cover data mining, so you can learn more about how you can advance your business and sell more of your products and services through data mining.

25 Data Mining Blogs

While doing online research on data mining techniques, you can find all kinds of scholarly articles on the subject. There are Web sites that define what data mining is, sites that go into detail about how it works and how different techniques may be used to get the information you want. There are even sites that discuss the social and political ramifications of data mining. But in today's world, information changes quickly and sometimes an article can be out of date only a few months after it was written.

This is when blogs come in handy. Blogs are updated regularly, and good bloggers will keep up with changes in information and technology almost as soon as they happen. Bloggers may even help readers who are new or unfamiliar with the subject to understand how things work better than any technical journal or website.
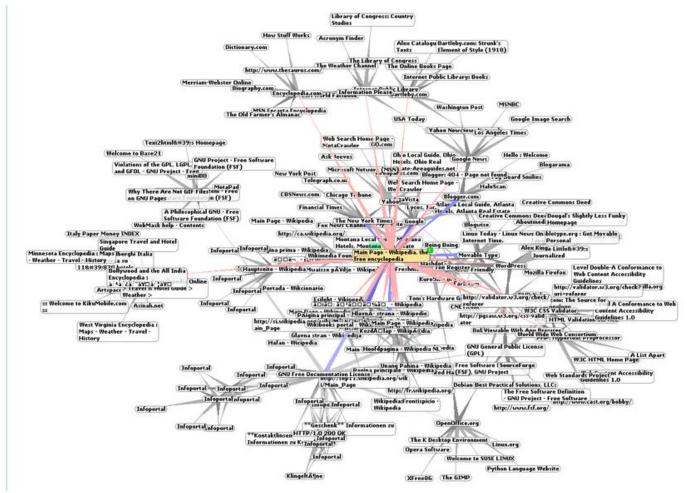
1. [K.D. Nuggets](#) A data mining blog on twitter, providing brief blurbs on current events and plenty of links.

2. [Data Mining: Text Mining, Visualization, and Social Media](#) A blog discussing the author's personal experiences with data mining.

3. [Occam's Razor by Avinash Kaushik](#) A market analyst shares his thoughts on data mining and web analytics.

4. [Geeking with Greg](#) A blog that seeks to examine the future of personalized information.

5. [CoolData](#) A blog looking at analytics and predictive modelling in relation to fund raising for higher education.

6. [Decision Stats](#) A blog discussing current trends in analytics and data mining.

7. [Fern Halper's data makes the world go 'round](#) A blog that focuses mostly on business and text analytics.

8. [Flowing Data](#) A blog that focuses on data visualization and statistics.

9. [Life Analytics](#) Subtitled "Practical Applications of Data Mining, Text Mining and Information Extraction,"this blog that looks at data mining's impact on society.

10. [LingPipe Blog](#) A blog that focuses on natural language processing and text analytics.

11. [Machine Learning (Theory)](#) A blog that focuses on machine learning and learning theory

12. [MarkTab Data Mining](#) A blog discussing how to use automated and semi-automated statistical techniques to make better decisions

13. [MineThatData](#) This blog provides a discussion on multichannel and database marketing.

14. [Neoformix](#) This blog focuses on discussions on how to discover and illustrate patterns in data, data as artform.

15. <u>Oracle Data Mining</u> This blog is dedicated to Oracle data mining, providing everything from technical details to people's opinions.

16. <u>Predictive Analytics World</u> A blog to promote Predictive Analytics World conferences, business-focused events for predictive analytics professionals, managers and commercial practitioners

17. <u>Revolution Computing Blog</u> This blog looks at news on statistics and the world of open source.

18. <u>Data Doghouse</u> This blog takes a look at performance management and data warehousing from both the technical and business ends of the question.

19. <u>Tim Manns' Data Mining Blog</u> A data-mining professional looks at terabyte data warehouses and the software's SPSS Clementine and SAS Enterprise Miner.

20. <u>Tom H. C. Anderson blog</u> A market analysts personal blog, discussing data and text mining.

21. <u>The Numerati</u> A blog which examines the rise of digital data and the applied mathematicians who used this data to make predictions on everything from business to voting and even other people's love lives.

22. <u>The Text Frontier</u> A blog which covers text mining, voice mining and a discussion of unstructured data analysis.

23. <u>No Free Hunch</u> This is the official Kaggle blog which discusses news, competition findings and other data-prediction related information. Kaggle is a platform for data prediction competitions that allows organizations to post their data and have it scrutinized by the world's best data scientists

24. <u>Tuned IT</u> This blog covers data mining and machine learning as well as AI research and implementation.

25. <u>Measuring Measures</u> This blog is dedicated to discussing how to gather data intelligence from information.

Tools for Data Mining

The image shown above is a data visualization of Wikipedia as part of the World Wide Web, demonstrating hyperlinks. Chris 73 created this image using TouchGraph GoogleBrowser V1.01, which is based on the related function of Google. This is just one tool in a global arsenal that designers and developers can use to formulate and forecast data.

But, if you're simply seeking a way to log all those business cards into a database, you might simply use Excel or, better yet, the open source speadsheet Calc, which is part of the OpenOffice suite. Students of both tools can learn how to extract information and compile new databases through both products. But, if you don't want that learning curve, you might try something simpler, such as Evernote.

The issue I have with Evernote is that it is impossible to analyse information from data, and difficult to transport to another data mining tool. Since data mining is all about analysing data, tools like Evernote just don't cut the cake. Extracting information is important as well, as you don't want to go through data input more than once if at all possible.

At the same time, you don't want to spend hours on data mining and analytics, as you probably have better things to do, and you probably cannot afford a person to do data input and analytics even on a part-time basis. But, if you have the money, time and help, you might look through the list of data visualization tools offered at Wikipedia. These tools are most closely aligned with visual perception, as they can create information graphics that can appeal to your sensibilities as well as help with client issues (TIP: You might be able to add the cost of any one of these products to the next client bill if that client expects analytics and forecasting.).

Fifteen Open Source CRM Solutions

CRM stands for Customer Relationship Management, and tools that offer this type of data input and analytics often cost and arm and a leg – suitable for corporations, but not for the lone eagle entrepreneur. However, some open source solutions exist, and – while not free – the cost is minimal for some spectacular results. That said, I am not endorsing any one of the tools listed below, as I have not tried them all. They are listed in alphabetical order to show my non-preference:

1. CentraView: Browser based Enterprise Java (J2EE) Contact Management, Groupware, Collaboration, CRM, SFA, Project Management software. Run locally or as a hosted service. Built on Apache Tomcat, JBoss, MySQL, Linux (Fedora, RedHat and others) & Windows.
2. CiviCRM: From its project website, it is described as a powerful contact, fundraising and eCRM system that allow you to record and manage information about your various constituents including volunteers, activists, donors, employees, clients, vendors, etc. CiviCRM is web-based, internationalized, and designed specifically to meet the needs of advocacy, non-profit and non-governmental groups. Integration with both Drupal and Joomla! content management systems. CiviCRM requires PHP5.2.x, PHP 5.3.x is supported starting with version 3.2.
3. Compiere: This company offers a complete ERP (Enterprise Resource Planning) and CRM application with support for marketing and sales, human resources, inventory control and more. There is even a full business-accounting package, something often seen in open-source software. Compiere is written in Java, JavaScript and PL/SQL, and it is compatible with JDBC and Oracle databases.
4. Concursive: This company is the developer of several products, including ConcourseSuite, a dedicated CRM solution, and ConcourseConnect, an integrated web-based software platform that brings together Social Business Software, Corporate Social Networking, Online Community, Business Directory, and CRM capabilities. The CRM is free up to five users, but you must host it yourself for that "price." Tools are written in Java and are compatible with MySQL databases.
5. Cream: This is a customer relationship management system for media organizations. The application tracks sales orders, payments, shipments, services, online and print subscriptions, and the effectiveness of promotional campaigns. Modules allow communication with customers via newsletters, email and a Web interface. Cream CRM runs on FreeDSB, Linux and Windows 2000/XP, and is written in Java and JavaScript.
6. Dolibarr ERP/CRM: Dolibarr ERP/CRM is a software built by modules addition used to manage small or medium companies, freelancers or foundations. This is an OpenSource project base on a WAMP, MAMP or LAMP server (Apache, Mysql, PHP for all Operating Systems).
7. Hipergate: This project's mission is to cover a full range of technical requirements in any organization. All applications are addresses from Internet Explorer, without needing any other additional software in the client computer. This suite is multi-company capable, and can be used in a single company, a corporate group or working as an ASP solution capable of serving an unlimited quantity of single customers. According to the Web site, this application is "completely free." Written in Java and JSP, hipergate is operating system-independent. Databases with which it is compatible include Microsoft SQL Server, MySQL, Oracle and PostgreSQL.
8. OpenCRX: openCRX is an open CRM solution that meets the needs of organizations requiring multifunctional, enterprise-wide coordination of sales generation, sales fulfillment, marketing and service activities to customers, partners, suppliers or intermediaries. openCRX features a Java API and a REST-service making available the complete API for third-party applications; you can also integrate openCRX with XML-RPC and there is even a specialized shop interface that makes it easy to use openCRX as a backend for shops; seamless integration with MS Exchange for contacts, calendars and e-mail is done with openCRX AirSync. openCRX runs on any platform with Apache Tomcat and one of the major database management systems (e.g. HSQLDB, MySQL, PostgreSQL, MS SQL, Oracle, or IBM DB2).
9. Open Taps: Open Source ERP + CRM is a fully integrated application suite that brings together top-tier open source projects to help you manage your business more effectively. And if you are a developer looking to customize opentaps for your organization or your clients, you might be happy

with their domain-driven architecture that embraces object-oriented development model with integrated Hibernate persistence and Spring framework.

10. SplendidCRM: This application is built on the Microsoft platform (Windows, ISS, SQL Server, C# and ASP). Designed for system integrators, SplendidCRM allows administrators to add user-customizable features such as .NET 2.0's Themes, Web Parts and AJAX.

11. SugarCRM: Despite my unwillingness to endorse products, I have noticed that many people who use CRM products seem very happy with Sugar. Over 6,000 customers and more than half a million users rely on SugarCRM to execute marketing programs, grow sales, retain customers and create custom business applications. SugarCRM runs on the leading cloud computing platforms, including Amazon EC2, Microsoft Azure, Sugar On-Demand and Private Clouds, offering customers unparalleled choice and control of their data and deployments.

12. Tustena: Every version of Tustena CRM is a complete solution designed to help you maximize your efforts. Their CRM solutions come in four different flavours: the "OnDemand", the "Business", the "Retail" and the "Business Source" (that includes the complete source codes and technical support by their team). Users who opt for the Open Source version get the source code and can modify it as they wish, relying on the community for support. Tustena CRM is written in C#, ASP.NET and JavaScript. It is compatible with Microsoft SQL Server.

    vtiger: vtiger CRM is a free, full-featured 100-percent open source CRM software ideal for small and medium businesses, with low-cost product support available to production users that need reliable support. Vtiger is built upon the LAMP/WAMP (Linux/Windows, Apache, MySQL and PHP) architecture, and also offers add-ons (Outlook Plug-in, Office Plug-in, Thunderbird Extension, Customer Self-service Portal and Web Forms) and support for other add-ons. vtiger is written in JavaScript, PHP and Visual Basic. It is compatible with ADOdb, MySQL and PostgreSQL databases.

13. XRMS CRM: This is a Web-based application suite that features business intelligence (BI) tools, Computer Telephony Integration (CTI), and advanced plugin architecture. CRM is operating-system independent, and it is written in an interpreted language (PHP). Compatible databases include ADOdb, SQL-based, Microsoft SQL Server, MySQL and other network-based DBMS.

14. Xtuple: The founders of the company that would become xTuple saw a compelling opportunity to build a new solution for small manufacturers – based upon open standards which would give growing companies the flexibility they needed. On the technological side, that meant leveraging all the work on robust open source building blocks such as the Linux operating system, the PostgreSQL database, and the Qt framework for C++.

15. ZOHO CRM: Outside the products, this site contains great information on how and why to use CRM management tools. Zoho CRM empowers organizations with a complete customer relationship lifecycle management software for managing organization-wide sales, narketing, customer support & service and inventory management in a single business system. Their free version is an "easy to use" hosted CRM for entrepreneurs and startups. Otherwise, Zoho CRM is a pay as you go (monthly or yearly) service so you can upgrade, downgrade or cancel at any time. Additionally, Zoho Projects provides REST style Web APIs over HTTP. With this API, users/developers can extend their web applications to push or pull data into or from Zoho Projects for application mashups. As the API is purely HTTP based, your application could be in any of the programming languages like Java, Python, ASP etc., to use the same.

Since I'm a designer/writer and not a developer, I wanted a simple CRM solution for a nonprofit that I work with. In the end, we chose a non-open source application, Salesforce, as they offer free licenses to certain nonprofits. Additionally, the Web-based interface seems easy to use. Additionally, when I surfed to the site, a sales rep was ready to talk with me via instant message to resolve my questions quickly. Let's say I was sold immediately. Now, to convince the board to "come on board" to expand our nonprofit's future reliability, credibility and organizational savvy.

Additional Data Tools

You may already be familiar with some other data mining tools already offered through DMXZone. They include the following:

Extensions

1. Advanced Data Calendar Bundle PHP: Add and edit events, appointments or birthdays in user-friendly inline popup window. Choose from Comma Separated (CSV), XML, Excel or PDF file formats and get a direct download. Requires DMXzone Calendar together with Advanced Tooltips and Advanced HTML Editor 2.
2. Feed Genie: Get your favourite feeds as Data Sources directly in Dreamweaver.
3. Feed Genie & HTML5 Image Enhancer Bundle: This bundle allows you to enhance images or feeds thumbnails with fantastic effects. Add your favourite feeds as data sources in Dreamweaver and create real-time animation effects, reflections, gradients, image filters and transition with only few clicks and without any Flash.
4. Universal Data Exporter PHP: Choose from Comma Separated (CSV), XML, Excel or PDF file formats and get a direct download.
5. Universal Data Exporter ASP: Ditto.

Notice that the above extensions with data, and they also offer ways to visualize that data. However, if you want to accumulate data on a Web site or other Web-based site such as Facebook, you may need to incorporate other tools.

TIP: Go to the results page for data mining search at DMX to find tutorials, premium content and more about data mining.

Other Web-based Tools

1. Facebook Insights: View and decipher Fan Page analytics from this useful tool. As I mentioned in the previous article, you can view demographics, increases or decreases in Page participation and more.
2. AWstats: Designed with flexibility in mind, AWStats can be run through a web browser CGI (common gateway interface) or directly from the operating system command line. Some Web hosting services offer this open source analytic tool as part of a hosting package.
3. Google Analytics: You can use the Web-based tool or download the plugin to use in WordPress. This tool offers insights into your website traffic and marketing effectiveness. As an afterthought, you might want to check out Google Apps for additional CRM materials that will integrate with any Google app site.
4. Open Web Analytics: Open Web Analytics (OWA) is open source web analytics software that you can use to track and analyze how people use your web sites and applications. OWA is licensed under GPL and provides web site owners and developers with easy ways to add web analytics to their sites using simple Javascript, PHP, or REST based APIs.
5. Webalizer: The Webalizer is a fast, free web server log file analysis program. It produces highly detailed, easily configurable usage reports in HTML format, for viewing with a standard web browser. Once again, this resource may be offered by hosting solutions without charge.
6. Woopra: Currently split into two services, a desktop client program and a web server application, Woopra radically decreases the load on network resources, harnessing the power of desktop processing to display graphic charts, graphics, and analysis. Woopra Plugins, which integrate Woopra's web analytics into website and blog databases and templates, generate even more information for the webmaster or blogger about their site's traffic and usage. Woopra offers a free plan and a variety of paid versions with additional benefits and features. For Windows, Linux and Mac.

I could go on, but the list of Web-based analytic tools is a long one. For a list of more Web-based analytic services, check the Wikipedia list. Also, you might want to read the article, Web-Based Analytics Vs Web Server Logs. Both analytics and Web server logs serve a similar purpose, but function differently and are intended for different objectives.

Conclusion

Designers and developers might have a similar interest in data mining, especially when it comes to client retention and organization. At that point, a CRM could come into play, so that you have easy access to client and resource information. When it comes to information about Web pages, especially a Web site, blog or your Facebook page, use some of the items listed above to learn more about your appeal across the globe.

Developers might stand to gain special interest in data mining and analytics when it comes to application development and customizing open source tools. This is a field that only will grow over time, as more entrepreneurs, small businesses, nonprofits and corporate entities seek to refine data mining on a competitive level. The ability to stay on top of resources with the touch of a finger is appealing, even to the lowest common denominator in the commercial/nonprofit sectors. Sometimes, it may be the only tool that these entities can use to stand ground in highly competitive fields.