

Tools of the Trade

Derek Willis

The Washington Post

SLA 2006

dwillis@gmail.com

This is a list of tools -- all freely available and cross-platform -- that may help you explore the world of information management and display within the newsroom. Most of them assume that you have the ability to install programs on your work computer. If you don't, don't despair - you can try these out at home. But be sure to read the docs!

Apache: The Apache web server is the most popular on the Internet, and is among the easiest to install and setup on Windows (Apple computers come with Apache built-in). The current version is 2.2.2, but version 2.0 is fine as well. Download it from <http://httpd.apache.org/>.

MySQL: MySQL is a powerful database server that has several advantages over MS Access or even SQL Server. The first is speed - MySQL sacrifices some of SQL Server's advanced features for more speed - and the second is its ease of use for Intranet deployment. One major difference is the way that MySQL treats dates: the natural syntax is YYYY-MM-DD rather than MM/DD/YYYY. Download from <http://www.mysql.com/>.

Perl/PHP/Python: These three are known as scripting languages, and are great utilities by themselves or in concert with other web applications. All work on any platform and come with Apple computers. My personal preference is Python, which I find easier to understand, but PHP is very common in Intranet applications. Download from <http://www.perl.org>, <http://www.php.net>, or <http://www.python.org/>.

Instiki: Instiki is wiki software that requires only one thing - the Ruby scripting language, which is also free. Instiki comes with its own built-in web server, so you don't even have to have Apache installed to run it (it runs on port 2500 so it won't conflict with Apache). Instiki's developers recently released a major upgrade that switched to a database storage unit, but you can use version 0.10, which does not require a database. Download from <http://www.instiki.org/>.

MediaWiki: MediaWiki is the software that runs Wikipedia and other wiki sites. It requires a database (MySQL is the default), along with Apache and the scripting language PHP. It has some more features that Instiki lacks,

but otherwise the two are very comparable. Download from <http://www.mediawiki.org/>.

xPDF: xPDF is a utility that can strip the text from an electronically-generated PDF document while preserving the layout, which can aid in getting PDF data into a spreadsheet or database. Originally a Linux/Unix utility, it also has a Windows version that works from the command prompt. Download at <http://www.foolabs.com/xpdf/>.

wget: wget is another utility that specializes in batch retrieval of files on the Internet. If you need to mirror a web site, fetch thousands of files from a directory or download only selected file-types, wget is a great solution. Like xPDF, wget is mainly a Linux/Unix program but Windows binaries are available and work via the command prompt. Download at <http://www.gnu.org/software/wget/>.