Digital terms for Journalists

The days are over when a journalist could ignore those geeks in the corner who typed lines of code, worked on the website and spoke in a bizarre language populated with acronyms. Any journalist's story now may be distributed with an API; information gathered by a reporter could be used in a mashup or shared via Scribd.

This glossary will help you wade through such terms. They relate to Web standards, programming, online tools, social networking, online advertising and basic technology. If you're particularly challenged, this thing called an iPhone is even defined for you. Get started now — the list will grow more quickly than your relatives join Facebook.

–Steve Myers

API (Application Programming Interface) — The way computer programs share data and functionality with other computer programs. APIs are an increasingly critical part of the Internet's interconnection. Many say that the future of the Internet lies in APIs because they help distribute and combine content. On the Web, APIs are generally special URLs that give back machine-readable data, in formats like JSON or XML, rather than human-readable data, which is usually HTML. Facebook, Twitter and Google Maps all have APIs that allow other websites or computer programs to use their underlying tools. The New York Times and NPR have also released APIs that allow other programs to draw on archives of movie reviews, restaurant reviews and articles.

algorithm — A set of instructions or procedures used in order to accomplish a task, such as creating search results in Google. In the context of search, algorithms are used to provide the most relevant results first based on those instructions.

Android — Usually used in the context of Android phone, Android is a free and open source operating system developed by Google that powers a variety of mobile phones from different manufacturers and carriers. It is a rival of the iPhone platform. In contrast to Apple's tightly controlled architecture and App Store, Android allows users to install apps from the Android Market and from other channels, such as directly from a developer's website — which allows for X-rated content, for example. Some well-known Android phones are the Nexus One, the Motorola Droid and HTC Evo. Expect to see competitors to the iPad running a version of Android.

app — Short for application, a program that runs inside another service. Many mobile phones allow apps to be downloaded, leading to a burgeoning economy for modestly priced software. Can also refer to a program or tool that can be used within a website. Apps generally are built using software toolkits provided by the underlying service, whether it is iPhone or Facebook.

AJAX — A bundle of technologies and techniques that allow a web page to do things quietly in the background without reloading the whole page. AJAX is not a programming language, but rather an acronym used to describe that bundle, "Asynchronous Javascript and XML." AJAX provides much of the functionality associated with Web 2.0. One of the first big uses of AJAX was Gmail, which allowed it to be much more responsive than other web e-mail at the time.

Atom — A syndication format for machine readable web feeds that is usually accessible via a URL. While it was created as an alternative to RSS (Real Simple Syndication) to improve upon RSS's deficiencies (such as ambiguities), it still is secondary to RSS. (See also, RSS)

blog — One of the first widespread web-native publishing formats, generally characterized by reverse chronological ordering, rapid response, linking, and robust commenting. While originally perceived to be light on reporting and heavy on commentary, a number of blogs are now thoroughly reported, and legacy media organizations have also launched various blogs. Originally short for "web log," blog is now an accepted word in Scrabble.

Blogger — A simple, free blogging platform created by Pyra Labs, which was sold to Google in 2003. It was one of the first mass blogging services and is credited with popularizing the format. Unlike WordPress, it is not open source. Many Blogger sites are hosted at blogspot.com.

civic media — An umbrella term describing media technologies that create a strong sense of engagement among residents through news and information. It is often used as a contrast to "citizen journalism" because it also encompasses mapping, wikis and databases. MIT has a Center for Future Civic Media.

cloud computing — An increasingly popular computing model in which information and software are provided on demand from over the Internet rather than staying on local computers. Cloud computing is appealing because companies can reduce the amount they spend on their own computer servers and software but can also quickly and easily expand as the company grows. Examples of cloud computing applications include Google Docs and Yahoo Mail. Amazon offers two cloud computing services: EC2, which many start-ups now use as a cheap way to launch their products, and S3, an online storage system many companies use for cheap storage.

client side — Referring to network software where work takes place on the user's computer, the client, rather than at the central computer, known as the server.

Advantages of doing so include speed and bandwidth. An example is Javascript, a programming language that allows developers to build interactivity into websites. The work is done within the browser, rather than at the hosting website. (See also server side)

CMS (Content Management System) — Software designed to organize large amounts of dynamic material for a website, usually consisting of at least templates and a database. It is generally synonymous with online publishing system. The material can include documents, photos or videos. While the first generation of content management systems were custom and proprietary, in recent years there has been a surge in free open-source systems such as Drupal, WordPress and Joomla. Content management systems are sometimes built custom from scratch with frameworks such as Ruby on Rails or Django.

CPA (Cost Per Action) — A pricing model in which the advertiser is charged for an ad based on how many users take a specific, pre-defined action–such as buying a product from an online store–based on viewing an ad. This is the "gold standard" for advertisers because it most directly matches the cost of an ad to its effectiveness. However, it's not commonly used since it's extremely difficult to measure: it is often unclear when or how to attribute an action to a specific ad. (Also sometimes referred to as Cost Per Acquisition.)

CPC (Cost Per Click) — A pricing model in which the advertiser is charged for an ad based on how many users click it. This is a common model for "search advertising" (the all-text ads associated with search results) and for text ads in general. CPC is well-suited for "directed" advertising, intended to prompt an immediate response, because a user's clicking on an ad shows engagement with it. Google AdWords is generally priced on a CPC basis.

CPM (Cost Per Mille) — Cost per one thousand (often views). Much of online advertising — particularly display advertising — is priced on a CPM basis. (Mille = Latin for one thousand; we use "K" for "kilo" almost everywhere else in tech, but "M" for "mille" here, which causes some confusion.) CPM is well suited for "brand" or "awareness" advertising, in which the primary purpose of the ad is not necessarily to prompt an immediate response.

Creative Commons — A flexible set of copyright licenses that allow content creators to specify which rights they reserve and which they waive regarding their work that is supposed to codify collaborative spirit of the Internet. There are six main Creative Commons licenses based on four conditions that creators can choose to apply: Attribution, Share Alike, Non-Commercial, and No Derivative Works. The least restrictive of the licenses is Attribution, which grants anyone, from an individual to a large company, the right to distribute, display, or otherwise make use of the work so long as the creator is credited. The most restrictive is Attribution Non-Commercial No Derivatives, which grants only redistribution. First released in December 2002 by the nonprofit Creative Commons organization, which was inspired by the open source GNU GPL license, the licenses are now used on an estimated 130 million works worldwide. The glossary you are reading is released under a Creative Commons Attribution Share Alike license in an effort to encourage wide distribution and contribution. (Also see open source)

CSS (Cascading Style Sheets) — Instructions used to describe the look and formatting for documents, usually HTML, so that the presentation is separate from the actual content of the document itself. If you watch a web page that loads slowly, you will often see the text first load and then "snap into place" with its look and feel. That look and feel is controlled by the CSS. CSS, which was first introduced by the World Wide Web Consortium in the late 1990s, helped eliminate the clumsy and often repetitive markup in the original HTML syntax. W3cschools.com has a great introduction to CSS with tutorials.

CSV (Comma-Separated Values) — An extremely simple data format which stores information in a text file. CSV is popular precisely because it can be easily read by many different applications, including spreadsheets, word processors, programming text editors and web browsers. Thus it is a common way for people, including governments, to make their data available. Each row of data is represented by a line of text. Each column is delimited/separated by a comma (,). To prevent confusion about commas in the data, the terms are often surrounded by double quotes ("). Many applications support the use of alternative column delimiters (the pipe character, |, is popular). Example below:
"Name","Address","email"
"Jack","1 Main St., Town, NY","jack@hill.com"
"Jill","2 Elm St., City, CA","jill@hill.com"

data visualization — A growing area of content creation in which information is represented graphically and often interactively. This can be used for subjects as diverse as an analysis of a speech by the president and the popularity of baby names over time. While it has deep roots in academia, data visualization has begun to emerge on content sites as a way to handle the masses of data that are being made public, often by government. There are many tools for data visualizations, including Seattle-based Tableau and IBM's Many Eyes. Data visualization should 1) tell a story, 2) allow users to ask their own questions and 3) start conversations.

document-oriented database — An increasingly popular type of database. In contrast to relational databases, which rigidly require information to be stored in pre-defined tables, document-oriented databases are more free-flowing and flexible. This is important when you don't know what is going to be thrown at you. Document-oriented databases retrieve information more

quickly, but store it less efficiently. The same document-oriented database might let you store the information for an article (headline, byline, data, content, miscellaneous) or for a photo (file, photographer, date, cutline). MongoDB is a popular open source document-oriented database.

Drupal — A popular content management system known for a vibrant open-source community that creates diverse and robust extensions. Drupal is very powerful, but it is somewhat difficult to use for simple tasks when compared to WordPress. Drupal provides options to create a static website, a multi-user blog, an Internet forum or a community website for user-generated content. It is written in PHP and distributed under the GPL open source license. Whitehouse.gov uses Drupal.

Django — A web framework that is popular among news and information sites, in part due to its origin at Lawrence Journal-World in Kansas. It is written in Python, a sophisticated dynamic language. Major projects built in Django include Disqus, Everyblock.com and TheOnion.com. News applications teams, including those at the Chicago Tribune and Los Angeles Times, use the framework to present large data sets online in easily accessible ways.

embed — A term meaning to place a specific piece of content from one web page inside of another one. This is often done using an embed code (a few lines of HTML and/or Javascript) that you can copy or paste. This is a common way for video content to be spread around the Internet and is increasingly being used for interactive components. A recent example is PBS Newshour's oil spill tracker widget, which was placed on many news sites around the country. Note: This is different from the newsroom sense of "embed," popularized during the 2003 Iraqi invasion, which means to have a journalist work from within a military unit.

EC2 — A computing power rental system by Amazon that has become popular among technology companies because it is much cheaper than maintaining your own computer servers. Users can host their applications on EC2 and pay depending on usage. EC2 is an example of cloud computing. (Also see cloud computing)

Facebook Connect — A technology from Facebook that allows a reader to log into a third-party website with their Facebook account, rather than creating a new profile for that website. Facebook Connect, which is an API, also allows the third parties to pull certain data from the user's profile, such as his or her name and age. In turn, the reader's activities on the website can also be displayed on her or his Facebook profile. Launched in 2007, Facebook Connect was one of the first examples of Facebook extending itself into a platform for the entire Web. (Also see OAuth, Open ID)

Facebook community page — Introduced in April 2010, community pages were created as a counterpart to "official fan pages," which are built around a specific person, company, organization, product, or brand. In large part, community pages are mostly auto-generated around interests or affiliations found in people's profiles, like cooking. There is not a way to actively add content to the page, unlike with Facebook groups. But because they are autogenerated, based on likes, they can quickly build gigantic memberships. Cooking, for example, has over 2 million fans. These pages are a bit confusing, and Facebook is still working on the kinks.

Facebook fan page — A Facebook profile for a specific person, product, company or organization, usually administered by official representatives. This is different from a Facebook personal page, which must be owned by an individual, and different from a Facebook community page, which is built around an interest not related to a brand, such as "cooking." It is also different from a Facebook group. Fan pages can gather thousands or millions of fans though "likes," and official posts by the page administrator generally go into the fans' news streams. Once a page has more than 25 fans, it can claim a short form URL, such as facebook.com/nytimes or facebook.com/wikileaks. Facebook community and fan pages are strong players in ongoing efforts to bring content to people where they already are, instead of requiring them to come to the content.

Facebook group — Facebook groups are analogous to offline clubs. Unlike Facebook fan pages, groups do not have to be administered by official representatives. In addition, the activity posted in groups does not get pushed into users' feeds. But as long as it has fewer than 5,000 members, Facebook groups are allowed to mass-message all their members.

Facebook personal page — A profile page tied to a single individual. What information is controlled (in theory) by the individual. However, because there is a 5,000-person limit to friends, some celebrities have fan pages instead. As of 2009, individuals can choose a username, which makes their page available at facebook.com/username.

Flash — A proprietary platform owned by Adobe Systems that allows for drag-and-drop animations, program interactivity, and dynamic displays for the Web. The language used, ActionScript, is owned by Adobe; this contrasts with many other popular programming languages that are open source. Creators must use Adobe's Creative Suite products and web surfers must install a Flash plug-in for their browser. Many claim that Flash players are unstable and inefficient, slowing down web pages and crashing operating systems. Apple has not allowed Adobe to create a Flash player for the iPhone operating system, which has created a feud between the two companies. HTML5 is emerging as an open alternative to Flash.

framework — A software package that makes writing programs easier by providing all the "plumbing" for a particular type of task (like writing a web app), allowing programmers to just "fill in the blanks" with their own project-specific needs. For instance, Web development frameworks like Ruby on Rails (written in Ruby, meaning programmers use Ruby to do the "fill in the blanks" tasks) and Django (written in Python), have easy-to-use, built-in support for common web development tasks, such as reading and writing to a database, writing content in html, and so forth. Watch Django and Ruby creators discuss the merits of their frameworks on DjangoProject.com.

Foursquare — One of many new mobile services, along with Gowalla, SCVNGR and others, that combines geolocation with game mechanics. Launched in 2009 at SXSW Interactive conference, Foursquare allows users to "check in" at locations (bars, restaurants, playgrounds and more) to inform people in their social networks of their whereabouts while earning badges, collecting points and becoming the "mayor" of certain locations. Despite a relatively modest user base at the beginning, Foursquare quickly attracted a lot of attention for its potential for marketing and customer brand loyalty.

geotag — A piece of information that goes with content and contains geographically based information. Commonly used on photo sites such as Flickr or in conjunction with user-generated content, to show where a photo, video or article came from. There has been some discussion of its increasing relevance with geographically connected social networking sites, such as Foursquare. Twitter has implemented geotagging, and Facebook has announced plans to do so.

Google AdSense — Google's online advertising network that allows content publishers to embed a piece of code to display Google ads on their sites. The ads are selected based on the content of the page. Ad revenue is split between Google and the publisher in an undisclosed proportion, generally believed to be two-thirds to the publisher. (Note: ads on Google's own sites are covered by Google AdWords, not AdSense.)

Google AdWords — Google's text-based flagship advertising product, which provides the lion's share of the company revenue. Ads are displayed on Google's own sites based on search terms that users type in, and advertisers pay only when the users click on them. The search terms, called keywords, are purchased by advertisers; availability of a given keyword is based in part on an auction system, and in part on the responsiveness of the audience.

Google Buzz — Launched in February 2010, Buzz is Google's attempt to counter Twitter and Facebook by leveraging the social graphs from users' e-mail accounts. A more sophisticated version of Gmail "status updates," Buzz allows users to post updates about what they are doing, link to what they are reading and post

their current locations. The service can integrate with other Google services, as well as feed into Twitter. Despite an initial burst of publicity, Google Buzz has not gained tremendous traction. It attracted criticism when Google automatically and publicly connected users with people they had e-mailed most often in the past, making private information unexpectedly available. Google released enhanced privacy controls after the controversy.

Google Docs — A free online service offered by Google, comprising word processing, spreadsheet, presentation and other software, all of which is "in the cloud." Users can work collaboratively on documents, editing them simultaneously. The service is increasingly being seen as eroding Microsoft Office's market share. The glossary you're reading right now was collaboratively created in Google Docs.

Google Wave — An online collaborative space introduced by Google in which people can communicate and work together in real time; it resembles a "souped up Instant Messenger." Participants can add rich text, images, attachments, videos and maps to create a multimedia collaboration. A playback option allows new users to get up to speed on projects and creates an environment that is both real-time and asynchronous. Despite a massive amount of attention, Google Wave has not gotten much traction. It is, as some people have said, "a technological solution in search of a problem."

HTML (Hypertext Markup Language) — The dominant formatting language used on the World Wide Web to publish text, images and other elements. Invented by Tim Berners Lee in the early 1990s, HTML uses pairs of opening and closing tags (also known as elements), such as <title> and </title>; each pair assigns meaning to the text that appears between them. HTML can be considered code, but it is not a programming language; it's a markup language, which is a separate beast. The latest standard of HTML is HTML5, which adds powerful interactive functionality.

HTML5 — The upcoming, powerful standard of Hypertext Markup Language, which has added advanced interactive features, such as allowing video to be embedded on a web page. It is gaining in popularity compared to proprietary standards, like Adobe Flash, because it is an open standard and does not require third-party plugins. Using HTML5 will allow web pages to work more like desktop applications. The latest releases of most browsers support HTML5 to varying degrees. HTML5 does not cover CSS and JavaScript, but often when people refer to HTML5, they often are using it as a blanket term, applying not only to changes to the HTML, but also to changes in CSS and JavaScript.

iframe — An HTML tag that allows for one web page to be wholly included inside another; it is a popular way to create embeddable interactive features. Iframes are usually constructed via JavaScript as a way around web

browsers' security features, which try to prevent JavaScript on one page from quickly talking to JavaScript on an external page. Many security breaches have been designed using iframes.

iPad — Released in April 2010, the iPad is Apple's tablet computing device, akin to a large iPod Touch; it uses the same operating system and development tools as the iPhone. It features a multitouch screen and comes in 3G and wifi versions. Some news organizations, including The New York Times, Wired and National Geographic, have created special applications designed for the iPad. Some have hoped that it would be the "Jesus" tablet that would breathe new life into legacy print publications. Upon its announcement in January 2010, many noted its name was reminiscent of feminine hygiene products.

iPhone — Apple's smart phone has sold more than 50 million units worldwide since it launched in 2007. The first smartphone to introduce multitouch screen capability, it is considered in the same vertical as the Blackberry, Google's Android and Palm Pre. The critical mass of iPhones, along with Apple's pre-existing iTunes infrastructure, allowed Apple to launch the first truly robust marketplace for mobile applications, creating a whole new microeconomy for innovation.

iPod Touch — Essentially an iPhone without the phone. Slimmer than the iPhone, the iPod touch can play music and run iPhone apps. It connects to the Internet via wifi.

JavaScript — A Web scripting language used to enhance websites; it can make them more interactive without requiring a browser plugin. JavaScript is interpreted by your browser instead of by a web server, otherwise known as a client-side scripting language. JavaScript files generally end in .js. Despite its name, it is not related to the Java language.

Joomla — A free, open-source content management built in PHP. It is more powerful than WordPress but not as powerful as Drupal. However it is known for its extensive design options. The name Joomla means "all together" in Swahili.

jQuery — A incredibly popular open source JavaScript library designed for manipulating HTML pages and handling events. Released in 2006, jQuery quickly gained widespread adoption because of its efficiency and elegance. The definitive feature of jQuery is its support for "chaining" operations together to simplify otherwise complicated tasks. It is the most popular JavaScript library.

JSON (JavaScript Object-Notation) — A Web data publishing format that is designed to be both easily human — and machine — readable. It is an alternative to XML that is more concise because, unlike XML, it is not a markup language that requires open and close tags.

Key/value store — A simpler way of storing data than a relational or document database. Key-value stores have a simple structure, matching values to accessible "keys," or indices. In Web development, key/value stores are often (though not always) used for optimization.

LAMP — An acronym referring to a bundle of free open-source Web technologies that have become incredibly popular as a method for building websites. The letters stand for the Linux operating system, Apache web server, MySQL database, and either PHP, Perl or Python. This is often referred to as a "LAMP stack." A rival alternative would be a bundle of Microsoft products. Serverwatch.com has a good explanation.

legacy media — An umbrella term to describe the centralized media institutions that were dominant during the second half of the 20th century, including — but not limited to — television, radio, newspapers and magazines, all which generally had a uni-directional distribution model. Sometimes "legacy media" is used interchangeably with "MSM," for "Mainstream Media." Legacy media sits in contrast with social media, where the production and sharing is of equal weight to the consumption.

library — In the context of programming, this contains code that can be accessed for software and Web development, enabling one to perform common tasks without writing new code every time. Many libraries are freely shared. One well-known library is jQuery, released in 2006 and now the most popular JavaScript library, which boasts that it allows coders to "write less, do more."

location-based services — A service, usually in a mobile Web or mobile device application, that uses your location in order to perform a certain task, such as finding nearby restaurants, giving you directions, or locating your friends. Foursquare and Gowalla are location-based services.

mashup — A combination of data from multiple sources, usually through the use of APIs. An example of a mashup would be an app that shows the locations of all the movie theaters in a particular town on a Google map. It is mashing up one data source (the addresses of movie theaters) with another data source (the geographic location of those addresses on a map).

metadata — Data about data. Examples of metadata include descriptors indicating when information was created, by whom and in what format. Metadata helps to organize information online and make it machine-readable. HTML is an example of metadata — it organizes the data in a web page so browsers can display it sensibly. Web pages often have hidden metadata that helps with their search engine ranks. Photos uploaded to Flickr carry metadata such as time taken, camera model and shutter speed. MP3s have

metadata such as the artist name, track title, album name and so on.

Microsoft Silverlight — Microsoft's answer to Adobe Flash, allowing the integration of multimedia, graphics, animations, and interactivity into web pages. It was initially released in 2007 and is occasionally spotted on the web.

mobile — An umbrella term in technology that was long synonymous with cellular phones but has since grown to encompass tablet computing (the iPad) and even netbooks. In retrospect, an early mobile technology was the pager. Sometimes the term is used interchangeably with "wireless." It generally refers to untethered computing devices that can access the Internet over radiofrequency waves, though sometimes also via wi-fi. Mobile technology usually demands a different set of standards — design and otherwise — than desktop computers, and has opened up an entirely new area for geo-aware applications.

MySQL — The dominant open-source database management system on the Internet. It is popular because it is a free and flexible alternative to expensive systems like Oracle. Projects that use MySQL include Facebook and Wikipedia. The SQL stands for "Structured Query Language" and "My" is the name of the inventor's daughter. It is officially pronounced My-S-Q-L, but you will often hear it referred to as "My Sequel." MySQL is a relational database management system, not a document-oriented database system. (Also see document-oriented database)

OAuth — A new method that allows users to share information stored on one site with another site. For example, some web-based Twitter clients will use OAuth to connect to your account, instead of requiring you to provide your password directly to that third-party site. It is similar to Facebook Connect. This allows sites to validate users' identities without having full access to their personal accounts.

ontology — A classification system with nodes or entities, that allows non-hierarchical relationships, in contrast to a taxonomy, which is hierarchical. Taxonomies and ontologies are important in content to help related articles or topics pages. (Also see taxonomy)

Open ID — An open standard that lets users log in to multiple web sites using the same identity through a third party. It is supported by numerous sites, including LiveJournal, Yahoo!, and WordPress. While Open ID has seen adoption among technical communities, its authentication method is not particularly intuitive, and it has not gained wide consumer acceptance.

open source — Open source refers to a philosophy and a means of developing and licensing software and other copyrighted works so that others are free to inspect, use and adapt the original source material. There are many open source licenses. Some licenses are considered permissive (e.g. MIT and BSD), allowing inclusion in proprietary works, while others (e.g. GNU GPL) require that the resulting derivative works remain under the same license if distributed. While the term originally stemmed from software practices, the concept has now been incorporated into other fields such as medicine and agriculture. Many of the most popular technologies used in content distribution, including languages and publishing platforms, are open source. The glossary you are reading was developed using open source methodology and is available under a Creative Commons license.

operating system — A basic layer of software that controls computer hardware, allowing other applications to be built on it. The most popular operating systems today for desktop computers are the various versions of Microsoft Windows, Mac OS X and the open-source Linux. Smart phones also have operating systems. The Palm Pre uses webOS, numerous phones use Google's Android operating system, and the iPhone uses iOS (formerly known as iPhone OS).

Palm Pre — A smart phone introduced in 2009 by Palm which uses webOS and allows for multitasking, unlike the iPhone. Despite rave reviews, the product is generally acknowledged to have come out too late to gain meaningful traction against the iPhone or Google's Android operating system. HP recently announced that it would acquire Palm, which was once the leading smart phone company.

peer-to-peer (P2P) — A network architecture in which users share resources on their own computers directly with others. Often used to speed up videos and large multimedia pieces that can take a long time to download. Napster was an early example of a popular use of peer-to-peer architecture, although it was not fully peer-to-peer. Today, Skype and BitTorrent are based on peer-to-peer technologies.

Perl — A dynamic language that is often used to parse and sort information because of its powerful abilities in manipulating text. Perl can be used to pull large quantities of data down from websites and standardize and replace information in batch. Perl was more popular in past years, especially in the computer-assisted reporting community, but it has been overtaken in popularity by languages such as Python and Ruby. Perl still has an active development community and is noted for the scope of its freely available libraries, which simplify development.

PHP — A popular web scripting language to generate web pages that was first developed in 1995, when it stood for "Personal Home Page." (It is now a recursive acronym, standing for "PHP: Hypertext Preprocessor.") Popular websites that are written in PHP are Wikipedia, Facebook and WordPress. It is criticized as being slow

because it generates web pages on request. However, Facebook recently released its internally developed version of HipHop for PHP, which is designed to make the language dramatically more efficient.

platform — In the technology world, platform refers to the hardware or software that other applications are built upon. Computing platforms include Windows PC and Macintosh. Mobile platforms include Android, iPhone and Palm's webOS. More recently, in an extension of its commonly used definition, Facebook has created a "platform," allowing developers to build applications on top of it.

Posterous — A blogging and publishing platform to which users can submit via e-mail. Through APIs, it can push the content to other sites such as Flickr, Twitter and YouTube. It is a for-profit company based in San Francisco that came out of the YCombinator seed start-up program.

PostgreSQL – An alternative to MySQL, another free and open-source relational database management system on the Internet. PostgreSQL is preferred by some in the technology community for its ability to operate as a spatial database, using PostGIS extensions. This enables developers to create applications that sort information based on geography, which can mean sorting by whether various places are within a certain county or pointing out the places that are geographically closest to the user.

programming language — A special type of language used to unambiguously instruct a computer how to perform tasks. Programming languages are used by software developers to create applications, including those for the web, for mobile phones, and for desktop operating systems. C, C++, Objective C, Java, JavaScript, Perl, PHP, Python and Ruby are examples of programming languages. HTML and XML are not programming languages, they are markup languages.

Python — A sophisticated computer language that is commonly used for Internet applications. Designed to be a very readable language, it is named after Monty Python. It first appeared in 1991 and was originally created by Guido van Rossum, a Dutch computer programmer who now works at Google. Python files generally end in .py.

relational database — A piece of software that stores data in a series of tables, with relationships defined between them. A news story might have columns for a headline, date, text and author, where author points to another table containing the author's first name, last name and email address. Information must be structured, but this allows for powerful queries. Examples include MySQL, Oracle, PostgreSQL and SQLite. Most modern websites use some kind of relational database to store content.

RSS (Really Simple Syndication) — A standard for websites to push their content to readers through Web formats to create regular updates through a "feed reader" or "RSS Reader." The symbol is generally a orange square with radiating white quarter circles. (Also see Atom)

Ruby — An increasingly popular programming language known for being powerful yet easy to write with. Originally introduced in 1995 by Yukihiro "Matz" Matsumoto, Ruby has gained increasing traction since 2005 because of the Ruby on Rails development framework, which can create websites quickly. Ruby is open source and is very popular for content-based sites.

Ruby on Rails — A popular Web framework based on the Ruby programming language that makes common development tasks easier "out of the box." The power of Ruby on Rails, which was developed by the Chicago-based firm 37 Signals, comes from how quickly it can be used to create a basic website.

S3 — An online storage system run by Amazon that's often used as a cheap way to store (and serve) photos and videos used on websites. It is short for Simple Storage Service. Its fees are often pennies per month per gigabyte, depending on location and bulk discount. The service is often used in conjunction with other Amazon Web Services, such as EC2, to allow customers to process large amounts of data with low capital investment. The New York Times used S3 with EC2 in this way to process its archives.

SaaS (Software as a Service) — A pricing strategy and business model, where companies build a software solution, usually business-to-business, and charge a fixed monthly rate to access it on the Internet. It is a type of cloud computing. Salesforce.com is the best example, but other notables include Mailchimp and even Amazon Web Services.

Scribd — A document-sharing site that is often described as a "YouTube for documents" because it allows other sites to embed its content. It allows people to upload files and others to download in various formats. Recently Scribd, which is based in San Francisco, moved from Flash-based technology to HTML5 standards.

scripting language — A programming language designed to be easy to use for everyday or administrative tasks. It may involve trade-offs such as sacrificing some performance for ease of programming. Popular scripting languages include PHP, Perl, Python and Ruby.

SEO (Search Engine Optimization) — A suite of techniques for improving how a website ranks on search engines such as Google. SEO is often divided into "white hat" techniques, which (to simplify) try to boost ranking by improving the quality of a website, and "black hat"

techniques, which try to trick search engines into thinking a page is of higher quality than it actually is. SEO can also refer to individuals and companies that offer to provide search engine optimization for websites.

SEM (Search Engine Marketing) — A type of marketing that involves raising a company or product's visibility in search engines by paying to have it appear in search results for a given word.

semantic web — A vision of the web that is almost entirely machine readable, in which documents are published in languages that are designed specifically for data. It was first articulated by Tim Berners-Lee in 2001. In many implementations, tags would identify the information, such as <ADDRESS> or <DATE>. While there has been progress toward this front, many say this vision remains largely unrealized.

server-side — Referring to when network software runs in a central location, the server, rather than on the user's computer, often known as the client. (Also see client side).

Sinatra — A lightweight framework written in Ruby that can be used to set up web services, APIs and small sites at lightning speed.

social graph — A mapping of the connections between people and the things they care about that could provide useful insights. The term originally promoted by Facebook and is now gaining broader usage.

social media — A broad term referring to the wide swath of content creation and consumption that is enabled by the many-to-many distributed infrastructure of the Internet. Unlike legacy media, where the audience is usually on the receiving end of content creation, social media generally allows three stages of interaction with content: 1) producing, 2) consuming and 3) sharing. Social media is incredibly broad and refers to blogging, wikis, video-sharing sites like YouTube, photo-sharing sites like Flickr and social networking sites like Facebook and Twitter.

structured thesaurus — A group of preferred terms created for editorial use to normalize and more effectively classify content. For example, the AP Stylebook is similar to (but includes more rules than) a structured thesaurus in that it gives writers preferred terms to use and standards to follow, so everyone following AP Style writes the word "website" the same way.

tag — A common type of metadata used to describe a piece of content that associates it with other content that has the same tag. Tags can be specific terms, people, locations, etc. used in the content it is describing, or more general terms that may not be explicitly stated, such as themes. The term "tag" is also used in the context of markup languages, such as <title> identifying

the name of the web page. In HTML, tags usually come in sets of open and closed, with the closed tag containing an extra slash ("/") inside. For example: <title>This is the Title.</title>.

taxonomy — A hierarchical classification system. In the world of content, this can be a hierarchy of terms (generally called nodes or entities) that are used to classify the category or subject content belongs to as well as terms that are included in the content. In many cases, website navigation systems appear taxonomical in that users narrow down from broad top-level categories to the granular feature they want to see. An ontology is similar to a taxonomy in that it is also a classification system with nodes or entities, but it is more complex and flexible because ontologies allow for non-hierarchical relationships. While in a taxonomy a node can be either a broader term or narrower term, in an ontology nodes can be related in any way.

Tumblr — A free short-form blogging platform that allows users to post images, video, links, quotes and audio. The company is based in New York City and competes with Posterous.

transparency — In the context of news and information, a term describing openness about information that has become increasingly popular. In many cases it is used to refer to the transparency of government releasing data to journalists and to the public. It is often used in the context of journalists being open about their reporting process and material by sharing with their readers before the final project emerges or providing more context in addition to the final product.

Twitter — A microblogging and social media service where users can send out messages limited to 14o characters. Launched in 2007, Twitter became popular in part because it had a set of APIs that allowed other developers to build tools on top if it. Twitter users came up with their own conventions, including the @ symbol to denote user names (@nytimes), and #, the hashtag, to denote subjects (#sxsw). Twitter computes Trending Topics, which give a real-time view into the most talked about topics on the service.

UI (User Interface) — The part of a software application or website that users see and interact with, which takes into account the visual design and the structure of the program. While graphic design is an element of user interface design, it is only a portion of the consideration.

URI (Uniform Resource Identifier) — The way to identify the location for something on the Internet. It is most familiarly in "http:" form, but also encompasses "ftp:" or "mailto:"

URL (Uniform Resource Locator) — Often used interchangeably with the "address" of a web page, such as http://hackshackers.com. All URLs are URIs, but not vice versa. While humans are familar with URLs as a

way to see web pages, computer programs often use URLs to pass each other machine-readable content, such as RSS feeds or Twitter information. In addition, words that appear in URLs often help boost search rankings, which is why many content sites are now shifting to URLs with headlines as opposed to data strings.

UX (User Experience) — Generally referring to the area of design that involves the holistic interaction a user has with a product or a service. It incorporates many disciplines, including engineering, graphic design, content creation and psychology. User interface is one element of user experience.

Web 2.0 — Referring to the generation of Internet technologies that allow for interactivity and collaboration on websites. In contrast to Web 1.0 (roughly the first decade of the World Wide Web) where static content was downloaded into the browser and read, Web 2.0 uses the Internet as the platform. Technologies such as Ajax, which allow for rapid communication between the browser and the web server, underlie many Web 2.0 sites. The term was popularized by a 2004 conference, held by O'Reilly Media and MediaLive, called Web 2.0. (Also see Ajax)

Web 3.0 — Sometimes used to refer to the semantic web. (Also see semantic web)

webOS — Operating system used on the latest generation of Palm smart phones, including the Pre and the Pixi. Apps for webOS are developed using web standards (HTML, Javascript and CSS), which means there is a low barrier to entry for web developers to create mobile apps for webOS as compared to other mobile platforms. It allows for having several applications open at the same time, unlike the current iPhone.

widget — In a web context, this refers to a portable application that can be embedded into a third-party site by cutting and pasting snippets of code. Common web widgets include a Twitter box that can sit on a blog, or a small Google Map that sits within an invitation. Desktop widgets, such as ones offered for the Macintosh Dashboard or by Yahoo!, can be placed on the desktop of a computer, such as for weather or stocks. Similarly, Android offers the ability to add widgets to the home screens.

wiki — A web site with pages that can be easily edited by visitors using their web browser, but generally now gaining acceptance as a prefix to mean "collaborative." Ward Cunningham created the first wiki, naming it WikiWikiWeb after the Hawaiian word for "quick." A wiki enables the audience to contribute to a knowledge base on a topic or share information within an organization, like a newsroom. The best-known wiki in existence is Wikipedia, which burst onto the scene around 2000 as one of the first examples of mass collaborative

information aggregation. Other sites that have been branded "wiki" include Wikinews, Wikitravel, and WikiLeaks (which was originally but is no longer a wiki).

WordPress — The most popular blogging software in use today, in large part because it is free and relatively powerful, yet easy to use. First released by Matt Mullenweg in 2003, WordPress attracts contributions from a large community of programmers and designers who give it additional functionality and visual themes. Sites that use WordPress include the New York Times blogs, CNN and the LOLCats network. It has been criticized for security flaws.

XML (Extensible Markup Language) — A set of rules for encoding documents and data that goes beyond HTML capacities. Whereas HTML is generally concerned with the semantic structure of documents, XML allows other information to be defined and passed such as <vehicle>, <make>, <model>, <year>, <color> for a car. It is the parent language of many XML-based languages such as RSS, Atom, and others. It gained further popularity with the emergence of Ajax as a way to send back data from web services, but has since lost ground to JSON, another data encoding format, which is seen as easier to work with.

Yahoo! Pipes — An online service from Yahoo! that provides a drag-and-drop visual interface to create interesting combinations of data. This is stuff you would otherwise need to know how to program to do. Instead, inputs, operators and chunks of logic are represented visually — as consoles connected by pipes — with information flowing from sources to output. It can import and out put in almost any common data format, including RSS, CSV, and JSON. Yahoo Pipes is an excellent resources for tech-minded, non-programming journalists.

Now More Context

We relate the different terms to each other and make sense of the whole thing.

Ruby, PHP and Python are all scripting languages that are commonly used in website development. PHP is used in WordPress and Drupal. Ruby is the basis for Ruby on Rails, a web framework. Python is the basis for Django.

Ruby on Rails and Django are similar and often considered rival systems because they both use a "Model, View, Controller" architecture. This means they are designed to separate the data (model) from the analysis of the data (controller) and its presentation to the user (view). Typically the model is stored in a database, the view is stored in html templates, and the controller is implemented directly in the underlying programming language (Ruby for Ruby on Rails, Python for Django).

XML and JSON are different ways of interchanging information. JSON is considered the better data exchange format whereas XML is considered a better document exchange format.

WordPress, Joomla and Drupal (listed in increasingly levels of complexity) are all content management systems written in PHP.

HTML5, which allows for more sophisticated use of graphics and videos in web pages, is considered an open standards rival to Flash, which is proprietary technology developed by Adobe Systems. Microsoft Silverlight is also a rival to Flash, but don't worry if you have never heard of it.

Facebook Connect, OAuth, and OpenID are different ways that users can use one account's information to log onto another website without having to create a new username and password.