

1

Wikis at Work

Wikis are websites that are collaboratively written by their readers. The software that makes wikis possible is called a *wiki engine*. This chapter introduces the wiki concept, and what you read here will apply to almost any wiki engine. The rest of the book, however, is devoted to one wiki engine in particular called MediaWiki, the wiki engine that runs what is arguably the world's most famous wiki, Wikipedia.

The idea that wikis are websites collaboratively written by their readers is simple enough, but the simplicity of the idea belies the profound impact a wiki can have on the flow of information among individuals. A wiki is to a typical website what a dialogue is to a monologue. On the surface, a conversation shares a lot in common with a lecture — in both cases, someone is talking and someone is listening, but the experience of a conversation is qualitatively different from the experience of either lecturing or being lectured, and the outcome of a conversation is qualitatively different from the outcome of a lecture as well.

In other words, authors are readers and readers are authors; there is no approval process required to post information on a wiki and there is no pre-ordained structure imposed on the content that is presented there. If you think of a regular website as a farm, with all the content organized into neat little rows of corn or beans, then a wiki is a meadow, teeming with grasses and wild flowers. A meadow isn't chaotic, however; there is order there, but it is a different kind of order. It's an emergent kind of order, one that evolves and is discovered, rather than imposed.

As with all definitions, this definition is only partly true. As time has passed, the principle of openness has been reshaped as a consequence of the hard realities of the world, and many wikis now restrict editing to certain users. Wikis have now become so popular that there are quite a few content management systems claiming wiki status with a completely different set of features than those conceived by the father of wikis, Ward Cunningham. He launched the first wiki (something he called a WikiWikiWeb back then) on March 25, 1995. A host of content management systems label themselves as wikis, even though they bear only a minor resemblance to the original wiki concept. This can make getting started with wikis a confusing affair.

Chapter 1: Wikis at Work

The Wiki Way is the name of a book by Ward Cunningham, and it is also a phrase used in reference to what was originally called Wiki Design Principles, which can be found on Ward's wiki at <http://c2.com/cgi/wiki?WikiDesignPrinciples>.

The most common question I am asked in my consulting practice goes something like this: “We have a content management system in place, but we’d like to have a wiki, too. How can my wiki integrate with the content management system?” This is like someone walking up to me and saying, “I have a pair of red shoes and a pair of black shoes and I’d like to integrate them into the same outfit.” I might suggest they wear both shoes — one red and one black. That might actually work if the only difference between the red shoes and the black shoes were the color. But what if the red shoes were running shoes and the black shoes were stiletto pumps? It would be very hard to get where you want to go.

The problem is this: A wiki is a content management system, not an alternative to a content management system. A website is a collection of related HTML pages that is accessible through the World Wide Web at a particular domain name (usually), and these pages are organized and linked to each other in a systematic way to make it easy for readers to find the information they seek. A content management system is a software application that provides tools to help people create and deploy websites. A wiki is a kind of content management system with a very special set of features that make it easy for people to use them to collaborate.

There are many different kinds of content management systems, each one suited to a different purpose, so the first question that really needs to be answered is “What are you trying to accomplish?”

There is a time and a season for everything. There is a time to wiki and a time not to wiki. This chapter aims to shed some light on when it makes sense to use a wiki, and when it may make sense to try a different approach. I start by exploring the history of wikis and why wikis have become such an item of interest in organizations. This is followed by a more detailed look at how wikis work and what kind of functionality is important when selecting a particular wiki engine. The chapter concludes with a discussion of best practices for running a successful wiki.

Once you know where you are going, it’s a lot easier to figure out what kind of shoes you need to wear in order to get there.

Wiki History

Ward’s original wiki is called the PortlandPatternRepository, and it can be found at <http://c2.com/>. The “WikiWikiWeb” name for the technology was inspired by the “Wiki Wiki” Chance RT-52 shuttle bus that runs between airport terminals in the Honolulu International Airport. *Wiki* is the Hawaiian word for quick, and that seemed to be an appropriate name for what Ward wanted to accomplish. His goal was to use the World Wide Web to develop a way for programmers to more readily share ideas about design patterns. In order for such a system to work, it needed to be something that was quick and easy to use.

In May of 1995, he invited a few of his colleagues to participate in his new site. It wasn’t long before the idea began to slowly catch on, and over the last decade a lot has changed.

Despite the success of PortlandPatternRepository, wikis did not start out as mainstream tools. It was the open-source software movement that first embraced the idea of wikis as an opportunity for widely distributed, decentralized teams of people to collaborate to produce software.

Chapter 1: Wikis at Work

The use of wikis by the open-source community is fitting, because wikis work on a principle similar to that which makes open-source software development so effective. The *Linus Principle*, named after Linus Torvalds, the original creator of Linux, is this: “Given enough eyeballs, all bugs are shallow.” The benefit of sharing the source code with your application is that you have more people who can look at the code and find problems. Wikis provide the same logic to content development: The more people who can both read and edit a document, the more likely it is that errors will be caught and fixed.

The technology behind a wiki is relatively simple: Ward’s contribution to humanity was not the code used to produce the first wiki (although I am sure it is fine code indeed), but getting the technology out of the way so that people can communicate and collaborate. In every organization on the globe, both large and small, there is information about the organization floating around inside people’s heads that needs to be documented in some way. Practitioners of knowledge management call this kind of knowledge, which is informal and learned largely from experience, *tacit knowledge*. What they call *explicit knowledge* is formalized and documented knowledge. The goal of knowledge management is to transform tacit knowledge into explicit knowledge: In other words, the goal is to get all that information floating around inside people’s heads written down. Much to everyone’s surprise, wikis have proven to be remarkably effective in this regard.

When I first learned about wikis in the nineties, I was skeptical because I immediately imagined the fun that malcontents would have defacing whatever site you tried to manage this way. It wasn’t until the success of Wikipedia that wikis caught my attention, along with the rest of the world. A relatively small team of technically savvy developers is one thing. A global pool of experts collaborating on an encyclopedia is another thing altogether. What is remarkable about Wikipedia is its scale and ambition. What is most surprising is that it works so effectively. This is what has gotten knowledge management experts so excited about wikis.

Wikipedia was an offshoot of the online encyclopedia *Nupedia.com*, founded by Jimmy Wales and funded by Bomis (something Wales has reportedly called “a guy-oriented search-engine”). Nupedia was founded in March of 2000 and established as a peer-reviewed encyclopedia with a seven-step editing process. This seven-step process proved to be rather cumbersome and not enough articles were being generated.

As a solution to this problem, Editor in Chief Larry Sanger proposed a “feeder” site to Nupedia based on wiki technology on January 10, 2001. The idea was that people could post articles on the wiki and after those articles had been properly vetted, they could be moved onto Nupedia. The use of a wiki would make it easier for users to contribute and, it was hoped, speed up the process. There was never any expectation at the time that Wikipedia would replace Nupedia, although that is what quickly happened.

Five days later, Wikipedia was formally launched, and within its first year it generated over 18,000 articles. By the time Nupedia closed up shop in September of 2003, Wikipedia boasted over 160,000 articles, written by volunteers. In three and a half years, Nupedia’s peer-reviewed process produced 24 articles, compared with 160,000 articles produced in the first year alone of Wikipedia’s existence. It is a remarkable example of the impact that moving from a centralized, formalized decision-making process to a more decentralized, informal process can have. More important, the quality of the content generated on Wikipedia was high, and the user base found it to be a very helpful research resource, so the traffic grew quickly.

Chapter 1: Wikis at Work

Web 2.0 and Social Media

While wikis are a distinct kind of website, they are often discussed along with other technologies under the label of *Web 2.0* or *social media*.

At one time, I considered “Web 2.0” to be a phrase in search of a meaning. My first reaction to the idea of Web 2.0 was a quick roll of the eyes. I fell into the camp shared by Tim Berners-Lee, inventor of what we can presumably call Web 1.0, who remarked that “Web 2.0 is, of course, a piece of jargon, nobody even knows what it means.” (IBM DeveloperWorks Interview, www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html).

Despite my dismissiveness, the Web 2.0 meme has proven its resilience and is alive and well, having evolved into a number of variant phrases, such as Andrew McAfee’s Enterprise 2.0, which he defines as “the emerging use of Web 2.0 technologies like blogs and wikis (both perfect examples of network IT) within the Intranet.” (Andrew McAfee, http://blog.hbs.edu/faculty/amcafee/index.php/faculty_amcafee_v3/the_three_trends_underlying_enterprise_20/).

J. Bonasia, writing for *Investor’s Business Daily* in June, 2007, said that users “are just getting familiar with the concept of Web 2.0, through which they can collaborate and share Internet content.” (*Investor’s Business Daily*, www.investors.com/editorial/IBDArticles.asp?artsec=16&artnum=3&issue=20070601).

According to Tim O’Reilly, Web 2.0 represents the movement to the “Internet as platform, and an attempt to understand the rules for success on that new platform” (O’Reilly.com, www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html). The term was coined in an effort to capture what was different about companies that survived the Internet bust of early 2000, and those that did not. As such, Web 2.0 is not a set of Web technologies per se; rather, it is a set of attributes shared by successful Internet companies.

The list of technologies commonly associated with Web 2.0 are wikis and weblogs, RSS, AJAX, Web services (SOAP, XML-RPC, ReST) and so on. Some of these are standards, some are concepts, some are architectures, and many of them have been around since the mid to late nineties. Unfortunately, the 2.0 designation implies new technology, although I do not think O’Reilly necessarily intended that.

While the definitions of Web 2.0 and social media are somewhat squishy, some common themes arise when pundits try to define them. Wikis and related tools all share four common attributes:

- ☐ Participatory
- ☐ Decentralized
- ☐ Linked
- ☐ Emergent

Wikis are clearly participatory. Unlike traditional content management systems, in which users have distinct roles and the set of content creators is entirely distinct from the set of content consumers (or, readers, as we once quaintly referred to them), all are equal (or mostly equal) in the public square of wikidom.

Likewise, wikis are decentralized in the sense that participants can be geographically disbursed. Wikipedia boasts authors from across the globe. They are also decentralized in the sense that wiki

Chapter 1: Wikis at Work

content isn't organized into a hierarchy and is not as structured as typical content managed by a content management system.

This decentralized content is structured by way of links, which can be old-fashioned hypertext links from one document to another document, or it can be a conceptual link made manifest by the sharing of a common tag (another word for what is essentially a keyword that represents the subject matter of a given page).

This participatory, decentralized, and linked collection of ideas is not organized in a top-down manner because there is no top or bottom. Rather, any order that arises is an emergent order. A system arises out of the interactions of many individual agents, each operating under its own set of rules, much like weather patterns emerge from billions of atoms acting the way that atoms do, completely unaware of the larger system in which they are unwitting participants.

New Business Models

In almost every case, these technologies, practices, and design patterns are continuations of the fundamental idea that has been central to the Web's success and pattern of development. Tim Berners-Lee did not invent hypertext. In his book *Weaving the Web* (HarperBusiness, 2000), he shares his experiences contacting commercial providers of hypertext systems in his efforts to convince them to open their platform. All the companies whose names remain obscure and unfamiliar refused to open their platforms. In the absence of cooperation from commercial vendors, Tim Berners-Lee developed the World Wide Web himself as an open standard.

Proprietary versus Open Standards

Prior to the Web, most businesses based their strategy on the creation of proprietary technology or platforms that provided them with a sustainable competitive advantage over any potential competitors. Whether you were a software developer or a content publisher, you differentiated yourself by being unique and maintaining strict control of intellectual property. The reason why commercial hypertext vendors resisted opening up their systems was because they feared that if they did, then they would not make any money from those systems.

They were right, of course, but they missed the point. While creating a hypertext system based on open standards may have made it impossible to make money selling that hypertext system, it also happened to create a platform through which money could be made (or efficiencies gained) that proved to be far more powerful than anyone could have predicted. The free and open nature of the platform made it ubiquitous and instantly relevant.

Network Effects

Network effects refers to the idea that networks become more valuable as the size of the network increases. Network effects tend to have a "winner take all" effect, with one platform emerging as the dominant platform. The World Wide Web itself is the perfect example. The more websites that were available on the Web, the more valuable the network became.

One way to differentiate social media from old media is to look at how network effects work in the old, proprietary environment and to contrast that with how they work now. I would argue that social media are harvesting network effects in a fundamentally new way.

Chapter 1: Wikis at Work

The best example of the old network effect is that of Microsoft Word. The reason I have a copy of Microsoft Word on my computer is because so many other people have a copy of Microsoft Word on their computer. The more people who use Microsoft Word, the more it makes sense for me to use Microsoft Word because it is simpler to read documents they send to me, and simpler for me to send them documents I've written, and so on. The only connection between Word documents is the fact that they share a common platform and that the readers of those documents need certain software applications in order to create or read them.

In the pre-Web days, the platform was a proprietary software application. Because Microsoft owned the platform, Microsoft made a lot of money selling licenses to use that platform. This is an example of a network effect in a proprietary environment, which happens to be a very favorable environment for the owner of the proprietary system who is able to succeed in this environment.

But what happens when you create an environment based on open standards, without any proprietary technology, such as the one created by Tim Berners-Lee? How does one company compete more effectively than another company and, most important, how is value captured?

The search engine Google is often mentioned as an example of a Web 2.0 company. None of the technologies mentioned earlier explain why Google is a Web 2.0 company. It has nothing to do with AJAX, blogs, wikis, or Web services. What Google has done is successfully capture network effects in an open environment.

Google's PageRank algorithm (which is how the search results are prioritized) is based in part upon how many other sites link to a given page. If you have two separate pages, both with similar content (as ascertained by word count and position), favor is given to the page to which more sites link than the other. Google infers that a page with more links to it must be better than a page with fewer links to it. Every day Google learns more about the content that is distributed on the Internet. By doing this, Google leverages the wisdom of the crowd, using the aggregate wisdom of Web participants to make more effective guesses about what specific content is most relevant to the searcher. As a result, Google's site makes it easier for me to find the information I am looking for.

Google is not the only company doing something like this. Flickr does the same for photos. Flickr takes a simpler and more direct approach by having visitors tag photos — it simply enables users to assign a keyword of choice to any photo they come across. This ad hoc system of keywords is called a *folksonomy*, a term used to differentiate this approach from a taxonomy because it is a decentralized approach to organizing content, as opposed to a *taxonomy* (such as the Dewey decimal system, or the Yahoo! directory) in which content is placed in taxonomic classifications by experts or specialists. Again, Flickr is leveraging the wisdom of the masses. The knowledge is a consequence of the steady aggregation of knowledge in the form of links created by human beings.

Flickr is better every time a user adds a tag. The value of Flickr isn't the repository of photos; there are plenty of sites for hosting photos, and the technology required to host photos verges on the trivial. However, as time passes, the database of knowledge about the photos increases, and the connections between photos that can be made by the folksonomic tagging of information means that Flickr increases in value every day and as a consequence derives true competitive advantage.

Google is continuing to build on this by offering new services. One such service enables Google users to create their own search engines. This is a wonderful service and I have used it to aggregate content on a variety of sites of interest to me. Moreover, when you are aggregating content, you are leaving behind a trail of knowledge and human judgment that Google will be able to use to make their site's

Chapter 1: Wikis at Work

search results even more effective. Google harvests human knowledge in creative ways, by interpreting the results.

By enabling users to create and edit articles, Wikipedia, too, is leveraging the collective wisdom of Wikipedia users. Much like Flickr does, Wikipedia (and all wikis run on MediaWiki software) enables users to add arbitrary tags (called *categories* in MediaWiki) that describe the content of the page. Wikipedia, too, is a better product every time a category is added to a page.

New Publishing Model

Content-oriented websites have moved in a clear progression from a proprietary model to something completely different, a path that is mirrored by the path from Nupedia to Wikipedia. This is a path that emerged from a scenario in which content creation is centralized and controlled and a clear distinction is made between author and audience.

When a new technology emerges, it is always thought of in terms of the technology it replaces. Wireless telegraphs, horseless carriages, and the computer desktop are all examples of understanding a new technology using the terminology of the past. Eventually, wireless telegraphs were called radios and horseless carriages were called cars.

This is the purpose of metaphor — to use one idea in place of another because it makes it easier to understand, or evokes a sense that would otherwise be difficult to understand using more accurate, yet abstract language. Hence the desktop metaphor. There aren't really documents sitting in folders on your computer, but it certainly helps to think of your files that way.

We use metaphors because they can help us understand complex ideas, and it provides a frame of reference so that we know how to think about something new in familiar terms. The problem with metaphors, however, is that they can be limiting. Thinking about content in terms of folders and documents has narrowed our view of what they actually represent and what can be done with them.

The old way of thinking about documents emphasized the resemblance of computer files to paper documents — that is, tangible, discrete, and permanent things that can be filed away in folders. If the desktop metaphor is the old metaphor, then what is the new metaphor? How are we to understand communication in the post-Internet world?

The first websites used an old publishing model. Before the Web democratized publishing, the publisher owned the platform (the printing press or the content management system) and the content, all of which was highly controlled by a select few. Now, sites like Wikipedia have turned this publishing model upside down, eliminating the difference between author and reader. On Wikipedia and similar social media sites, authors are readers and readers are authors.

Wikis, Blogs, and Meme Trackers

Word documents are based on the metaphor of the typewritten page. Early websites were modeled after traditional print publications, and e-mail was delivered just like its printed, enveloped, and stamped counterpart. These tools are now officially out of fashion.

When a business adopts social media tools such as wikis and blogs within the organization, they are using the tools as a replacement to older forms of communication, such as e-mail. In doing so, they are abandoning the transitional tools based on older technology, and embracing new tools that leverage the power of the new technology.

Chapter 1: Wikis at Work

The one thing that social media sites, wikis, blogs, and meme trackers share in common is an understanding of the fundamentally dynamic nature of information. Content evolves. It is shared, modified, and shared again. It changes over time, it appears in different forms.

In the old way of thinking, if something has been “documented,” the implication is that the information contained therein is correct, complete, authoritative, and permanent. No such assumptions are made with blogs and wikis. In the wiki application MediaWiki (the software that runs Wikipedia), an article is defined as a collection of revisions. There are no definitive or authoritative articles on MediaWiki; there’s only the most recent revision.

Time plays a pivotal role in both. Blogs are organized according to when they were posted and, secondarily, by category or topic. A blogger doesn’t revise earlier posts; if a correction needs to be made or if new information surfaces, then a new post is all that is required. Much like on a wiki, everything on a blog is provisional.

It is the very public and transparent nature of wikis and blogs that creates their value. E-mail is fundamentally a private form of communication. The content goes from one mailbox to the next and the pieces of information in them remain separate from each other, discrete little bits of data hidden away in folders, much like the paper documents that serve as the underlying metaphor. Communicating and collaborating with wikis and blogs opens up that process and creates opportunities to discover new things, to make connections between things that we might not have thought of or understood before.

The network effect isn’t driven by the format because the format is open; the network effect is driven by the participants themselves and their aggregate wisdom. Web 1.0 is the linking of one HTML page to another. Web 2.0 is ferreting meaning and creating value through emergent properties associated with aggregating human judgment. Organizations that do a better job of making information available in useful formats will succeed, whereas those that attempt to control their information with proprietary constraints will whither.

In the post-proprietary world, the nodes of the network are not connected simply by sharing a common platform. In the post-proprietary world, the nodes of the network are points of data, information and ideas that are linked and aggregated and universally available.

Tim Berners-Lee and others originally envisioned the Web as a global repository of human knowledge, but the Web is not a library or a warehouse. As it turns out, the Web is emerging as a source of discovery, a phenomenon that, like other phenomena, can be analyzed and studied empirically and from which inferences can be drawn with a scope and a scale unknown before.

The Web does not simply store knowledge; it creates it.

Now, organizations both large and small are adopting wiki technology for a variety of purposes. Teams of developers still use wikis for documentation and project management. Some companies use wikis as the engine that powers their intranet — it’s free to install and easy to learn, so no other content management system offers a better cost/benefit ratio. MediaWiki is the software that runs Wikipedia. Because of Wikipedia’s success, MediaWiki is one of the most commonly used wiki engines available. It’s open source and free and runs on PHP and MySQL, making it easy for many organizations to adopt it. MediaWiki is not your only option, however, so in the next section I will go into greater detail about MediaWiki’s features and how they compare with more traditional content management systems.

Chapter 1: Wikis at Work

Web Content Management Systems

Content management systems are software applications used to facilitate the creation, storage, and distribution of digital content, and a wiki is a kind of content management system, with a twist. There are three areas where wiki engines really differentiate themselves from other run-of-the-mill content management systems: *access control*, *content authoring*, and *site organization*. In the following sections, I'll look at each one individually and discuss the features in depth.

Content Management Life Cycle

A Web content management system is a software application that provides tools to support the various activities required to maintain a website. As content flows through a content management system, it flows through four distinct phases, and all content management systems provide tools for each of these phases.

Content Acquisition

All content management systems receive their content from somewhere, a process I refer to as *acquisition*. The content may come from a legacy system or from another website in the form of syndicated content, or the content can be generated and edited directly in the content management system itself. Content management systems usually provide some kind of interface that enables users to create content, often through a Web browser. They also provide a system for managing workflow, which tracks content through various stages — from authoring to editing and ultimately to being published on the site.

Content Organization

Content management systems organize content so that readers can more easily discover the information they are looking for. This organization is called *information architecture*. Users find information on a website in two ways: they browse the site, navigating from page to page looking for the information, or they search for the information using a search engine.

Typically, content management systems provide a means for organizing content into a hierarchy, which is reflected in the system of navigation through which a user browses a site by following links. They also provide some form of search based on keywords, or some more elaborate scheme.

Content Storage

Content management systems also provide storage for content. How this is done varies from system to system. In some cases, the content is stored in a relational database; in others it is stored as XML on the file system. Many early wikis stored content simply as wikitext in plain text files. MediaWiki, too, stores content as raw wikitext, but it stores it in a database. It also happens to store all the previous versions of every page, which is a useful feature for websites that anybody can edit. From time to time, you'll find the need to roll back to earlier versions of a page.

Content Distribution

Finally, content management systems provide a means of distributing the content, which in the case of Web content management means providing a system of dynamically generating pages, and a set of tools that enable the publisher to schedule when content is viewable and by whom it is viewable.

Chapter 1: Wikis at Work

Wikis are unique in how they shepherd content through these stages. In the following sections, you will learn more specifics about the wiki approach to workflow, content authoring, and site organization, and the specific features offered by MediaWiki to support these processes.

Workflow and User Management

Because wikis are websites that are collaboratively written by their readers, the most unique characteristics of a wiki can be found in the systems that support user access control and workflow.

User Access Control

There are three stages to access control:

1. **Authentication:** This is the stage during which the system becomes reasonably assured that the person accessing the site is who they say they are. Most content management systems, including MediaWiki, do this with passwords.
2. **Authorization:** The authorization aspect of user access control works by assigning users roles. The role to which a user is assigned determines to which content objects that user has access. This can limit what a user sees on the site, as well as what a user can do.
3. **Activity tracking:** The final step is activity tracking, which means that the system generates an audit trail so that you can determine who did what to your site.

You will see this access control system at work with Wikipedia. If you have not registered on the site you have the role of an anonymous user, which means that you can view articles, but you cannot create them. In order to create articles, you need to be a registered user. Once you have registered for the site, you can customize certain features, such as the skin being used; you can create your own user page; and you can both create your own articles and edit articles that already exist.

Workflow Policies

Whereas access control systems control who can perform what task on a given content item, workflow takes this a step further and enforces a set of policies based on the state of the document. Workflow represents what tasks are to be performed, in what order they should be performed, and who should perform them in any given stage. From a content management perspective, workflow is the approval process as content moves from the authoring phases to the publishing phase of its life cycle.

In a typical workflow system, a user who is assigned an “author” role can create an article and submit it for approval by someone in an “editor” role. Once the editor approves the content, it will be published. In the world of wikis, there is no distinction between the two. Any changes made to an article are immediately published.

Control over the content is exerted by the fact that changes to articles are tracked, so that one can easily find out who made a particular change. From the access control perspective, most of the management comes from the ability to track what was done, rather than to use authorization to limit what one can do. This is why logging plays such an important role in wiki management.

Chapter 1: Wikis at Work

Change Monitoring

As one might expect, one layer of defense is to simply monitor changes that have been made to the wiki. In addition to monitoring changes, you want to be able to do something about fixing, or editing, unwanted changes, such as rolling them back to a previous version. Therefore, the “change monitoring” approach requires two basic features: the ability to monitor recent changes, plus some kind of version control.

Recent changes can be monitored as follows:

- ☐ Most wikis have a Recent Changes page that lists all the pages that have been changed and who made each change.
- ☐ E-mail notification of changes is just an e-mail version of the Recent Changes page, but with the convenience of notification.
- ☐ A variant of e-mail notification is support for RSS syndication, which means you can monitor a wiki for recent changes using your favorite RSS reader.
- ☐ MediaWiki allows you to differentiate trivial changes from more substantive ones. For example, you may not want to be notified by e-mail every time someone fixes a spelling error.
- ☐ If more than one person has been tasked with monitoring changes, another useful feature tracks whether a recently changed page has been checked yet, reducing the possibility of duplicating work. On MediaWiki, this is called marking a page as *patrolled*.

Version Control

I once encountered a philosophical debate about whether wikis should have version control. The idealist in the conversation argued that version control was against the “wiki way” and somehow lacked philosophical purity. The realist argued that people make mistakes and sometimes deliberately do bad things, so the ability to roll back changes was indeed a good thing and a feature that all wikis should have. I’m pleased to report that the realist won the argument in the broader marketplace of ideas, and many (if not most) versions of wiki software have version control.

Features include the following:

- ☐ The ability to roll back changes to the previous version
- ☐ The ability to compare different versions side-by-side
- ☐ The use of diffs between versions so that specific differences between them can be easily identified

Spam Prevention

Another approach is to monitor the content of changes programmatically, and this is sometimes referred to as spam prevention. This differs from user access control in the sense that it monitors wiki edits based on the content of the edit or the patterns of user behavior. Systems can block access to IP addresses and URLs, or they can block the posting of individual changes based on the following:

- ☐ Maintaining a spam blacklist, restricting access from certain domains
- ☐ Restricting the use of certain words or phrases, using word lists or regular expressions
- ☐ Blocking access based on excessive activity

Chapter 1: Wikis at Work

- ☐ Blocking by IP address or name
- ☐ Blocking content by type (or size)

Content Authoring

When we read a Microsoft Word document on our computer, we think nothing of the fact that not only can we view the document, but we can edit it as well. When dealing with online content, the fact that we can directly edit the content we are viewing is something of a novelty because in most cases, the content we encounter is read-only. There is, in fact, an effort to separate the creation of content from the design of content in the underlying technology (think HTML and CSS), and many websites have a publication mentality that draws a clear distinction between the readers and writers (hence the term *read/write web* that is used to refer to tools like wikis).

The wiki approach to authoring shortens the distance between editing and publishing a page in two ways. First, you edit the page using the same application used to view the page — a Web browser. Second, edits are posted immediately. There is no staging of draft versions, and no workflow requirements.

- ☐ Wiki pages are editable through a Web interface so that no special software is needed, other than a Web browser.
- ☐ Users with access to the site can edit pages directly, and the changes are published immediately.
- ☐ Wikis use a special markup sometimes called *wikitext* to specify formatting on the text of the page, or to automatically create links to other pages (see Chapter 4 for more information about wikitext). Many wikis now boast a WYSIWYG (what you see is what you get) interface because many users are more comfortable writing pages this way. Most of these interfaces are embedded into Web pages, but many require the use of a specific browser, such as Firefox or Internet Explorer. Ideally, a wiki should make editing available anywhere, on any browser, which is why some form of wikitext is required.

Organization

Ultimately, the goal of a content management system is to organize content in such a way that people can find it when they need it. The way in which content is organized depends on the goals of the site and the nature of the content itself.

All content management systems serve as a repository for content, but these systems are more than just a repository for the same reason that a library is more than just a repository for books. Documents in a content management system are organized in the same way that books in a library are organized. In a library, books are grouped together by a classification system so that like subjects are located in one place. For example, biographies are in one section, fiction is in another, and so on. There is also an index, which shows the exact location of any book within the library.

When you go to the library, you might decide to browse the books, rather than go straight to the index. In this case, you could walk to the section you are interested in and begin looking at the spines of the books to see if you can find something you like. Conversely, you might go to the library with a different goal in mind. That is, if you want to find a specific book, then

Chapter 1: Wikis at Work

you will go to the index. (When I was a child, this index was a physical card catalog, long rows of wooden boxes containing 3×5 index cards. This has since been converted to a digital index in those libraries that can afford it.) These two activities, browsing and searching, are also possible with content management systems. In fact, you can think of a library as a content management system, but one that deals with physical content, rather than digital.

Taxonomy

Most sites organize their pages by grouping similar pages together, the way the library groups similar books together. This classification is called a *taxonomy*, and on most sites it manifests itself as a hierarchical taxonomy, with a home page, sections, and subsections. For example, a typical newspaper website might have the hierarchical organization shown in Figure 1-1.

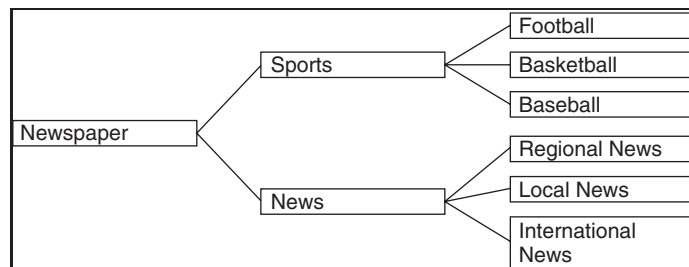


Figure 1-1: Typical hierarchy

The content of the site is first classified as either news or sports. Then, each of these classifications is subclassified, so that news is broken down into local news, international news, and regional news, and sports is broken down into football, basketball, and baseball.

As you navigate through the hierarchy from the home page, to sports, to baseball, you are arriving at more narrowly defined categories. The purpose of this kind of hierarchy is primarily to assist the user when navigating the site.

In a well-designed site, the taxonomy ultimately translates into site navigation and there should be a correlation between the taxonomy and the URL. In the example shown in Figure 1-1, the URL for basketball could be `http://choate.info/Sports/Basketball`.

One of the advantages to organizing sites in this way is that it makes it easier for users to guess the URLs for different areas of the site. After seeing the `http://choate.info/Sports/Basketball` URL, it would not be such a big leap to suppose that if you wanted to read something about baseball, you could go to `http://choate.info/Sports/Baseball`. Likewise, if you just wanted to see what kind of sports you could find more information about, you could navigate to `http://choate.info/Sports`. Having an intuitive URL namespace greatly improves the quality of a site.

These hierarchical structures can be topic-based, like the example I just used, but content can be grouped according to a variety of different criteria. Weblogs organize content in reverse chronological order, rather than by topic (some weblogs support categorization of content, but the basic structure of the weblog is chronological, as it is modeled after a journal or diary).

Chapter 1: Wikis at Work

Wikis are organized using a flat hierarchical structure. One consequence of this is that URLs are simple. For example, the following URLs link to Wikipedia's articles about football, basketball, and sports in general. All article URLs are this simple:

```
http://en.wikipedia.org/wiki/Football
http://en.wikipedia.org/wiki/Sport
http://en.wikipedia.org/wiki/Basketball
```

Folksonomy

Imagine if the library let you put the books back anywhere you wanted — that would lead to chaos, right? This means that if you put the book back into a different section, then other people wouldn't know where to find it because it wasn't put away according to the system. That would be bad in a library, but not in a wiki. Wikis employ a system of organization that lets you “put the book” wherever you please.

The reason is very simple. Wikis don't like hierarchies. They employ what is called a *folksonomy*. In terms of the URL space of a wiki, the hierarchy is flat. There are no sections and subsections. All pages maintain simple, non-nested URLs. Additional layers of organization evolve as tag-based folksonomies.

The excessively rural-sounding folksonomy is not the study of senior citizens; nor is it a measure of how much money grandpa has — it's just a clever way to refer to taxonomies that are created by a site's users in a decentralized, ad hoc manner. It is accomplished by allowing each user to assign tags to pages. Tags are really nothing more than searchable keywords that users decide to apply to a page.

In the case of MediaWiki, users can set categories for pages. There is no predetermined list of categories, and the terms used are entirely up to the user. This is in stark contrast to a taxonomy, which uses controlled vocabularies and rigidly defined structures into which individual units of content must be organized.

A list of categories on Wikipedia can be found at <http://en.wikipedia.org/wiki/Wikipedia:Browse>.

When to Wiki

I've seen business people and educators throw wikis at many problems (as the latest cure-all), only to see them splat against a wall and slowly slide down into a puddle of ooze. Wikis are wonderful things, but only when used correctly. The successful operation of a wiki requires both the right kind of technology and the right kind of governance. You cannot load up MediaWiki, flip a switch, and expect a wiki to perform wonders for you. You need to apply a wiki to the right problem, and you need to manage it properly in order to derive the greatest benefit. In this section, you will learn about the key management elements that contribute to successful wiki implementations.

There are two ways in which organizations begin to use a wiki. The first is by way of a top-down decision that takes place when someone in senior management decides that a wiki is the solution for some particular problem and thus mandates its use. The second way is by way of a grass-roots movement whereby individuals or workgroups begin to use wikis because they help them get their jobs done.

The very open and decentralized nature of wikis makes the grass-roots path the most common way that wikis find their way into an organization. Providing an environment in which wikis are allowed to

Chapter 1: Wikis at Work

emerge in a grass-roots fashion has some definite advantages. In the following sections, you will learn how best to foster the successful use of wikis within your organization. It is something of a contradiction to mandate the use of wikis. A better approach is to persuade through success. First, provide a fertile field in which the wiki seed can take root and thrive.

Running a Successful Wiki

Despite declarations of “Web 2.0” and the read-write Web and other trendy nomenclatures, the rules for a successful wiki are very similar to the rules that one should apply to any community site (that’s what we used to call them before we started to call them social media sites). One could argue that a wiki is a modified forum, as it retains many forum features.

The advice contained in this section is based in large part on my years of experience managing community-oriented sites for which the users are the main contributors of content, much like wikis. I also owe a large debt to Christian Wagner and Ann Majchrzak, whose paper “Enabling Customer-Centricity Using Wikis and the Wiki Way,” provided me with some particularly useful insights (see the *Journal of Management Information Systems*, Winter 2006–2007, Vol. 23, No. 3, pp. 17–43).

Their research focused on enhancing constructive customer engagement in a wiki, which is not necessarily how most companies will use one. However, the principles they suggest are good ones that, in my experience, do in fact foster a sense of community and collaboration.

In their review, they compared wikis operated by the *Los Angeles Times* (thinly veiled with the name “Boomtown Times”), Novell, and, of course, Wikipedia. The *Los Angeles Times* wiki was a dismal failure, the Novell wiki was a moderate success, and the Wikipedia wiki was, and is, of course, a smashing success.

Wagner and Majchrzak offer six propositions based on their research, but I have taken the liberty of condensing them into four rules of thumb, based upon my own experience with managing collaborative websites.

Alignment of Goals

Wikis got their start by being used by programmers to document software projects. This is an ideal use for a wiki because there is a strong incentive on the part of the programmer to participate. A social contract is at work: The software needs to be documented and all participants must be kept in the loop, so each programmer keeps his or her documentation updated with the understanding that the other programmers will do the same, in a mutual back-scratching arrangement.

This apparent no-brainer is, apparently, not a no-brainer. At least, it is not a no-brainer to the editors of the *Los Angeles Times*, which decided that it would be good to have a “wikitorial” — an editorial composed and edited by the masses. The *Los Angeles Times* started by posting their editorial, and then provided a wiki for the public to respond. As Wikipedia has learned, there is no alignment of goals among political types, and the partisans reigned supreme during the very brief life of the wikitorial. In the organic, evolving world of wikis, consider the wikitorial an evolutionary dead end.

At first, the users made a good faith effort to collaborate on an editorial, but they soon concluded that producing a single editorial that was acceptable to everyone was not going to happen, and there had already been attempts to delete the entire editorial, so by the second day, they forked the editorial so that it would be possible to represent different points of view. Once news of the wikitorial

Chapter 1: Wikis at Work

experiment showed up on Slashdot, a technology-related news website (<http://slashdot.org>), it attracted a lot of attention and was soon followed by pornographic posts, and so on. On the third day, the wikitorial was shut down.

The kind of vandalism encountered by the *Los Angeles Times* represents the nightmare scenario that is almost always raised as an objection to using a wiki. In fact, the first time I saw a wiki I thought it was a lousy idea for this very reason. As it turns out, this kind of defacement is not as common as one might think; and when a wiki is set up and managed properly, that kind of mischief can largely be avoided.

In the late nineties, I was responsible for what we called *community publishing* sites. It wasn't a wiki per se, but it had many of the features of a wiki, the most important one being that we (the newspaper) used the Internet to let the community participate in the publishing process. The site was called NCHome-team.com, and it represented a partnership between *The News & Observer* and WRAL-TV5 in Raleigh, North Carolina. It was a statewide high school sports site. Coaches and interested parents were recruited statewide to update the rosters each season, and then to update scores after the games on Friday night (which often finished too late to make it into the paper and/or there was not enough space).

Following were the main concerns we had when launching the site:

- ☐ Would the sites be as credible as the newspaper itself given the fact that the content was published without being vetted by an editor?
- ☐ Would the coaches reliably post their scores?
- ☐ Would the coaches post inaccurate results?
- ☐ What kind of liability would the newspaper have as a publisher? If acting as a publisher (as it does when things are printed), then the newspaper is responsible for all the content that is published. That means if someone is libeled in the newspaper, then the newspaper is responsible. (The first reaction was to avoid any activity that would make the newspaper look like a publisher — in other words, it didn't monitor posts.)

We quickly learned that the coaches were enthusiastic participants, and they were just as committed as we were to making sure that the information was timely and accurate.

In retrospect, it's easy to see why fears about false sports scores being posted were unfounded. While coaches did have an incentive to win, they had no incentive to cheat and post false scores because with so many other people at the game, they would easily get caught. The transparency of the process meant that it was in everybody's interest to post factual data. In other words, the goals of the entire community were aligned.

The reason why the *Los Angeles Times* wikitorial failed is because an editorial is a point of view about a controversial subject. The goals of the individuals on either side of the debate are to discredit those who disagree with them and to establish their worldview as pre-eminent. In other words, the goals of the left and the right are not aligned. Therefore, I recommend no bipartisan wikis. Ever. There is no such thing.

This does not mean that everyone who participates in a wiki has the same goals, or that they involve themselves in the same activities. Goal alignment only means that their goals are not in conflict; they all head in basically the same direction.

Chapter 1: Wikis at Work

A Culture of Collaboration

In practice, the most common problem encountered by new wikis is that it can be difficult to get people to participate. Several psychological and organizational barriers need to be overcome. Most important, in addition to needing the technical apparatus to operate a wiki, you also need an organization with a culture that fosters collaboration. If you don't, your wiki is unlikely to thrive.

For example, a certain government agency has decided to launch a wiki that will capture all of the undocumented but highly useful information that floats around in people's heads. They are facing two sources of internal resistance.

First, they have a hierarchical culture whereby every communication is approved by proper channels. Being propositioned by some young twenty-something about brain-dumping your wisdom into a wiki after having every utterance scrutinized by your superiors for your entire career is like suddenly being told by your wife of twenty years that she thinks you should loosen up a little and get a girlfriend.

The second source of resistance is the fact (or perception) that once your brain is dumped, it becomes communal property; and while you may be fairly certain that you will continue to get your paycheck, you are not so certain that you will continue to get credit for your faithful fidelity and the cultured wisdom you have nurtured for so many years. Owning information is a source of power; that's why it can be so hard to get people to share it.

In one case, a department was more than willing to post content in a content management system, but they were unwilling to do the same work if it was with a wiki. The reason? They feared a lack of control. It can represent a loss of ownership for people. When people are rewarded for individual output, they are going to be less inclined to participate in a project with collective output. While they may not say this aloud, they are worried about whether they will still get credit for their good ideas and hard work.

Universities are also experimenting with wikis. If you were a professor, you might reason that because Wikipedia is such a wild success, it would be fantastic to set up a wiki for the class. Then, instead of requiring students to write papers for an audience of one (which is you), they can write them for their peers, their fellow students; and for posterity, all the students that will follow. In that case, a classroom becomes a source or repository of knowledge. You even dreamily fantasize about students correcting and expanding upon the postings of other students in a communal editing effort whereby everyone is both student and professor.

The only problem is that students have no interest in correcting (or updating or expanding) another student's work. What's in it for them? They annoy a potential date and don't really have much to show for their work. There are two reasons to go to college: to learn and to acquire documentation that you have learned in the form of a transcript or diploma. I've used papers that I have written in graduate school as part of my portfolio. What kind of portfolio do I have if all the work was done in a wiki?

Most of your school life is spent being told to do your own work and keep your eyes on your own paper, while being forced to read wordy honor codes and the like. Traditionally, schools have not fostered a collaborative environment, so students aren't quite sure how wikis fit in with the culture of the school. In fact, wikis are an excellent tool for the classroom, but you need to be prepared to help the students unlearn some of what they have learned about what is appropriate behavior in school, just like employees of the government agency have to relearn what's appropriate for them.

Chapter 1: Wikis at Work

The one common theme that runs through all of these examples is that the goals of the participants were not aligned. In some cases, their respective goals were in direct opposition to each other, while in other cases there was a belief that participating in the wiki would not provide enough individual benefit. If you want to derive a benefit from collaboration, you need to ensure that everybody in the organization also benefits from collaboration.

Community Custodianship

I have already mentioned that when I first worked on community publishing sites for a newspaper, we conscientiously avoided creating the appearance that we were the “editors” of the content. In this case, we were doing so in order to avoid liability for what was posted on the community sites. This meant that we didn’t actively monitor user posts and that we wouldn’t remove posts unless a member of the community raised a concern with us.

What we had done inadvertently was to shift the monitoring responsibility to the community itself. Again, this was not for any altruistic reason, such as a belief in decentralized decision-making. As it turns out, however, letting community sites be managed, in effect, by the community is an important component of successful sites.

Despite the open nature of wikis, an effective wiki is not an egalitarian free-for-all. Just as the members of the community share reading and authoring privileges, they must also share custodianship of the community. The community rules the community. In this custodial role, the community of users needs to establish rules of conduct for contributors to the site, and they need to monitor user activity, to ensure that it is in conformance. The custodial role means that users are not only responsible for identifying suspect content, but they also serve on the decision-making bodies that establish guidelines regarding when such content is deleted, or when users should be banned.

Clearly Defined Rules for Posting Content

Successful custodianship means that in order to get your users to participate fully, your wiki needs to have clearly defined rules and processes. These rules include a clear description of the kind of content that should be contributed to the wiki as well as rules for handling disputes. For example, Wikipedia has “five pillars” that define the character of Wikipedia. The following is a sampling of a few of the rules:

“Wikipedia is an encyclopedia incorporating elements of general encyclopedias, specialized encyclopedias, and almanacs. All articles must follow our no original research policy and strive for accuracy; Wikipedia is not the place to insert personal opinions, experiences, or arguments. Furthermore, Wikipedia is not an indiscriminate collection of information. Wikipedia is not a trivia collection, a soapbox, a vanity publisher, an experiment in anarchy or democracy, or a web directory. Nor is Wikipedia a dictionary, a newspaper, or a collection of source documents; these kinds of content should be contributed to sister projects, here, Wiktionary, Wikinews, and Wikisource, respectively.”

“Wikipedia has a neutral point of view, which means we strive for articles that advocate no single point of view. Sometimes this requires representing multiple points of view; presenting each point of view accurately; providing context for any given point of view, so that readers understand whose view the point represents; and presenting no one point of view as “the truth” or “the best view.” It means citing verifiable, authoritative sources whenever possible, especially on controversial topics. When a conflict arises as to which version is the most neutral, declare a cool-down period and tag the article as disputed; hammer out details on the talk page and follow dispute resolution.”

Chapter 1: Wikis at Work

The rules are very explicit and leave little room for ambiguity. The preceding rules and others can be found at the following locations:

- ❑ http://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines
- ❑ http://en.wikipedia.org/wiki/Wikipedia:Five_pillars

In addition to establishing rules, you need to seed your wiki with content when it is first launched. The presence of content will facilitate the creation of even more content. One of the advantages of seeding the wiki prior to opening it up to a larger group is that the pages that you create serve as a kind of template for the new users to refer to when creating their own pages. In other words, they serve as an example of the kind of content you want to see on the site; and, it is hoped, having seen an example, people will be more comfortable producing their own content for the site.

Monitoring User Behavior

When Ronald Reagan talked about nuclear arms reduction treaties with the former Soviet Union, he espoused the following philosophy: “Trust, but verify.” Running a wiki requires trust on the part of management in the capacity of their employees, their customers, and the community at large to behave reasonably well, most of the time. Because it is not realistic to believe that they will behave reasonably well all of the time, then you must switch to “verification” mode and monitor behavior.

Despite the part of the definition declaring that wikis are sites that anybody can edit, the truth of the matter is that if you let just anybody edit it and do not, at the same time, provide a mechanism for proper oversight, your wiki will not work.

The ability to monitor user behavior creates transparency, and transparency is good. The very fact that behavior can be monitored will keep most of the behavior that needs to be monitored from ever happening. In fact, while the most common objection managers have to using wikis is fear of vandalism, the biggest problem they end up having is just the opposite: no activity at all.

The monitoring requirement varies according to how widely available the wiki is. In other words, a workgroup wiki behind the corporate firewall needs less monitoring than a customer-accessible wiki that the public can see.

It is also important that monitoring and policing the wiki remain the responsibility of the community. As I said earlier, community custodianship is one of those factors that creates well-run wikis, and one of the roles the community plays while acting in the capacity of custodians is the role of monitor. Not only should the community itself be the monitor, it should also be the body that helps to determine what the rules are in the first place.

Monitoring behavior can be more than simply a policing role. As mentioned earlier, one of the reasons employees can be reluctant to participate is a fear of losing credit. If anybody can edit a document, how am I going to get credit for writing this one? Most wikis can now track changes (MediaWiki can), and you can monitor activity on the wiki as a means of identifying good uses of wikis. This is especially true in educational settings where students might be graded on their activity.

Chapter 1: Wikis at Work

Wikis in the Enterprise

There are a lot of wikis on the market, both open source and commercial. If you want to learn more about the others, Wikipedia is a good place to start:

- ❑ http://en.wikipedia.org/wiki/Comparison_of_wiki_farms
- ❑ http://en.wikipedia.org/wiki/Comparison_of_wiki_software

Summary

In this chapter, you learned what wikis are, you learned about the role of a wiki within an organization relative to other content management systems, and you discovered some rules of thumb for managing a successful wiki. In the next chapter, the discussion moves away from a general discussion about wikis to a more specific discussion about MediaWiki. In Chapter 2, you will learn how to install and run MediaWiki software, including system requirements, options, and alternatives.