Multilingual Websites with Open Source Content Management Systems

Todd Kelsey, Illinois Institute of Technology

Abstract

Open source content management systems can be installed free of charge on an entity's servers or web hosting account, and provide a way for organizations to develop multilingual web sites. There are some challenges in finding a good open source content management system, but there are several that can fit a variety of needs. An example CMS is Plone, which has strong support for different languages, and which also integrates tools for managing the translation of content.

Introduction

The purpose of this article is to introduce the concept of an open source content management system that can support multiple languages, and to relate a journey of exploration that allowed the author to create a sample article in ten different languages, using the Plone system.

What is Open Source?

The term “open source” basically means software that can be downloaded and used free of charge. The creators of software not only provide the files, but also allow anyone who is interested to download the “source code”, which allows another person to more easily improve upon or otherwise add something to the original program. What has happened in effect is that an entire global community has sprung up, where teams of mostly volunteer programmers have joined forces to create new and innovate software, which provides a free alternative to commercial software.

Content Management Systems

The purpose of a content management system is generally to make it easier for an individual or organization to manage online content.

For example, the majority of websites are created manually, where each page that you see was created by a program such as Dreamweaver, and when something needs to be changed, someone needs to find the file that corresponds to a given page, make the change, and upload it to a *server*.

(When you make a web site, you ultimately need a web hosting account, at a company like www.hostway.com. The web hosting company maintains a set of specialized

computers called servers, which are connected to the Internet continuously, so when a person views your website, they are downloading the web page from the server to their own computer, using an internet browser such as Internet Explorer or Firefox. You can set up your own web server, and come companies and educational institutions have their own web servers, but most entities end up getting a separate web hosting account.)

So when your website gets bigger, maintaining it can become more challenging, especially if you have multiple people working on it, if there are different versions of articles or content that you need to keep track of, and especially if you're thinking about making a multilingual website, where you have the desire to make some or all of the content available in different languages.

What a content management system does is to allow you to concentrate on the content, essentially. In a sense, a content management system is like having a person maintaining your website for you. You create the content, and the content management system, or CMS, creates the appropriate files, places them in the right spot on the server, and keeps track of them in case you ever need to make changes. CMSes are especially helpful for organizations and websites where you want to have a variety of people all contributing content to the website. All the participant needs to do is to go to the Internet, log on to the CMS, and add their content, by typing it in, uploading a document, etc.. Whoever is administering the CMS can log on, create a new user, and allow new participants to add, change, or delete content.

One of the most helpful and useful things about a CMS is the ability to set permissions, where you can set a desired amount of access to different parts of the site, so some visitors may be able to read the content, and others may be able to submit, but only specified people might be allowed to delete, for example.

Another nice thing about a CMS is that you can involve people in making a website without requiring them to be technically proficient. That is, typically a web page is created by a web developer, who understands all the related technical issues and is often the person who is "uploading" the content to a web server. So in many situations, certain people create the content, and then pass it along to a web developer, who is the one who puts it up there. What a CMS does it provide a way for anyone with an Internet connection to contribute to a given site. It can certainly help to have a developer still involved, especially if there may be the need to customization, so a CMS doesn't affect the job security of web developers. In fact, web developers may like CMSes, because in certain cases, instead of having to upload and maintain a site, they can help participants to have direct access to content, so that the developer can concentrate on other things, such as customization.

Multilingual Websites

It used to be that multilingual websites were only within the grasp of organizations and corporations with large budgets. There are enough challenges with making a website, let alone trying to have versions of the content in different languages. But increasingly, more

websites are offering content in different languages, and I believe that open source content management systems may open the door for some entities with limited resources.

To be sure, to make a website in different languages, or to offer content in different languages, there's no reason you have to have a content management system. You could create the original language version of a website, and then, with careful file management, you could have the text of the site translated, make a copy of the files, and have a link to another version of the site. Alternatively, it may be possible to translate some of your content and make it available as a download in PDF format (For more info on PDF is, see http://www.adobe.com/products/acrobat/adobepdf.html).

But sooner or later, you may come to appreciate some of the things that a content management system can do, including allowing you to create user accounts so that anyone with an Internet connection, anywhere in the world, can easily add or modify content on the site. And when changes need to be made, the ability to keep track of particular versions of content on the site, known as *version control*, is an important feature.

With multilingual websites, not only will you have different versions of content in a specific language (such as the original, final, latest version, etc.), but a particular version of an article might be translated into different languages. This is where a content management system could end up being your best friend.

Translation Management

In this article, I wanted cover some basic concepts, and then relate how I personally ended up looking at content management systems. The thing that interests me in CMSes in particular is that some of them have built-in *translation management* – the ability to help a person keep track of different language versions of content. For the research that I am interested in doing for my PhD dissertation, I knew that I couldn't afford a commercial content management system, and I was excited to learn that there were open source CMSes that had some of this ability.

What excited me the most is the idea that I could create multilingual content, without necessarily having to delve into the complexities of how to incorporate all the different kinds of language characters on a Web page. And it was also exciting when I realized that an open source cms could potentially allow me to create the native language version of an article on the site, and then give direct access to a human translator, so that a new language version of an article could be fairly easily added.

A Note About Machine Translation vs. Human Translation

Machine translation is a process where computer software translates content from one language to another, with varying degrees of success. In general, you get what you pay for. If you use computer software to translate material, you may be able to save time and money, but the accuracy of translation will come nowhere near a human translating it. It's

not unheard of for a company or organization to "try" translation software, and then to end up hiring a professional translator in the end to "fix" things, but this could end up costing more than it would have cost originally to translate content.

My own personal recommendation is, if you're thinking about creating and releasing content in different languages, use human translation, whether paid or volunteer, and whenever possible, have the translation actually reviewed by another person. One area where a content management system can help is by allowing a native speaker to log in to your website, and have a "managed" environment to create a translation of a particular article. This can make it easier and more realistic to work with "volunteer" translation, or in the case of a corporate website, with bi-lingual speakers.

In the end, there's no real substitute for professional, human translation, but volunteer translation seems like at least a second best.

(Note: Even though machine translation is still in its infancy, a man named Franz Ochs has made some significant advances at Google Labs, using a lot of processing power, and a technique whereby human translated texts are analyzed, to "teach" the computers to translate better. So while good machine translation is not really available at the moment, don't be surprised if the folks at Google make it more available. For more information, keep your eyes on: http://googleresearch.blogspot.com/2006/04/statistical-machine-translation-live.html)

Case Study: Plone

I am not an authority on content management systems, but I concluded that the best way for me to relate the information I've learned for this presentation is to recount some of my own explorations.

One of the lessons that you quickly learn about open source software is that as with translation, in some ways, you get what you pay for. In other words, the software may be free, but it can take an investment of time to determine what you need, and to get things running. Still, for me, I didn't have any alternative, so I just began looking into things, drawn by the vision of a website where visitors could click on little flags, to see versions of content in their own native language – the dream of making content accessible to a global audience.

I began by developing a list of target languages that I ultimately wanted to develop content for, and this is one of the considerations that has an impact on what content management system will serve your needs. Most if not all content management systems have the ability to allow a person to create links between pages, and if all you need to do is create a bilingual English/Spanish site, for example, you might not even need to use a "multilingual" CMS per se – this is because some languages are easier to display on the Internet than others.

But I wanted to imagine what it would be like to think of a "top ten" list of languages, which might cover a significant portion of a global Internet audience. My list isn't meant to be definitive, and it is somewhat arbitrary, but I ended up wanting to have the ability to display and manage content in: English, French (which covers a significant part of Africa as well as France and other countries), German, Spanish, Portuguese, Russian, Chinese, Japanese, Korean and Arabic.

It is more challenging to display more "complex" characters such as Chinese and Japanese, so I was looking for an open source content management system that could handle these languages. It was also important to me to be able to handle Arabic – and this was a cutoff point for some CMSes. There is a technical standard called UTF-8, which is basically a standard that allows for the display of more complex characters, so if a CMS supports UTF-8, then it opens the door to a number of languages. However, Arabic reads right-to-left, and that brings a new twist to the picture.

Downloading and Installing

The nature of content management systems is that at some point, someone is going to have to download and install the specialized software on a server. For me, I had some technical ability, but I realized that it may be a bit over my head to do all of the installation myself, so my goal was to find a solution where I wouldn't necessarily have to install it. The goal of this article isn't to try and cover the technical intricacies of downloading and installing – but be aware that at some point, someone is going to have to do it. This may involve working with a developer, or as I show below, sometimes the web hosting company can set up the CMS for you. And when the CMS is up and running, most or even all of the subsequent development of your site can be done from simply logging in to the system from an internet browser, and using the automated features of the system. In other words, in many cases, once you get past the hurdle of installing the system, you may be home free.

Technospeak: For those of you who are technically inclined, basically most open source CMSes require a web server running software such as PHP, etc. – and in many cases, your existing web hosting account may support a CMS. In most cases, you end up needing to download a tarball or compressed archive of some kind, unpacking it on the server, and configuring some kind of database connection, such as MySQL. Also, in some cases, you can install a CMS on your own personal computer, to try it out "locally".

Comparison Shopping

I discovered a pair of helpful sites that allow the CMS explorer to take a closer look at a content management system before trying it out in earnest:

CMS Matrix (www.cmsmatrix.org) – This site allows you to select various CMSes, and generate a feature comparison, so you can see if a given CMS has the particular features you are looking for.

Open Source CMS (www.opensourcecms.com) – This site allows you to actually try out many CMSes, without having to install them yourself. It's a great tool for taking a closer look at things without having to make arrangements for downloading a CMS and getting it installed somewhere.

Open Source Host (www.opensourcehost.com) – The great thing about this place, and some other Web hosting providers out there, is that they offer an option for a "push button" installation of some of the popular content management systems. Open source host goes very far in this regard, offering very reasonable prices, and installing the content management system for free. So for something like 8.95$ USD monthly, you could get started with a content management system, and not necessarily have to hire a developer, or be all that technically inclined.

What Led Me To Plone (www.plone.org)

There are a number of different multilingual capable CMSes out there (you can do a comparative check on CMS matrix, just as I did), but my requirements and discussions with various people led me to try out Plone. What caught my attention with Plone was that it supports right-to-left language display, such as Arabic, and this made it a winner in my eyes.

What you start realizing when you look into open source CMSes is that they have varying levels of features, and varying levels of support in a developer community. What this means is, that in the open source world, the number of developers who like a particular product, in part determines how good the product is. A group of 3 or 4 friends might get together and make a content management system, or a brilliant individual developer might create something, but there are limits to what a small group of developers can do on a volunteer basis, and the thing that makes an open source project shine is the growth of its developer community.

So the ideal combination in some cases is that a for profit entity, such as a sponsoring corporation, provides some funds to a foundation, which is connected to an open source project. And then you would have a good developer community, with people from around the world contributing. These considerations weren't immediately apparent to me, but I started realizing that it's a bit of a Wild Wild West in the open source CMS world, with quite a large number of projects out there. And when you start looking closer, you can see some signs of how much support there is, such as how many people are involved, how many features there are in a given CMS (by comparing ones at www.cmsmatrix.org for example).

So I am not claiming that Plone is "the best open source CMS", but it has a good feature set, it seems to have fairly deep support for multiple languages, and it was the first CMS I came across where I was able to test it and achieve my goal of taking a simple article, translating it into my ten languages, and having each of these versions accessible through language selection flags.

Plone has its advantages, and its disadvantages. It seems to be fairly powerful, but I found that it didn't seem to have many companies where you could order a web hosting account and get Plone pre-installed. I ended up getting an account at a company called Zettai, in the Netherlands (www.zettai.net), and I'm working on learning enough so that I can get Plone installed on my own server. I've also had conversations with some developers who feel that Plone is very processor-intensive; this means that it can use of a lot of power on the web hosting server computer. For small websites with small audiences, this is no problem. But for the global website creator with dreams of large audiences dancing in their head, the challenge is that when you have a lot of people "hitting" a website and it is processor-intensive, the site may respond slowly. With Plone, there are fixes, such as having a fast server, and other technical tweaks which can be done. But I mention it to say that Plone is not perfect.

However, at the risk of getting this paper rejected for transgressing the boundaries of academic decency, I do have to say that Plone was my first love. I ordered the hosting account, and started exploring the features, and found a "product" in Plone called LinguaPlone.

LinguaPlone – CMS Add-ins

Usually a multilingual-capable CMS doesn't start out that way – someone comes along and develops an "add-in" that extends the ability of the CMS to include multiple languages. In the case of Plone, the LinguaPlone add in brings in a feature that caused me to fall in love with the system. And the LinguaPlone add-in comes with Plone, you simply need to "turn it on", so to speak, when you have Plone up and running.

Basically what LinguaPlone allows you to do is to define what languages you want to "enable" on your website, and when you start enabling different languages, you begin to realize how much power is lurking beneath the surface of a content management system, especially if you've ever worked with traditional ways of making websites, where you have to manually create pages and keep track of them.

I was overjoyed to find that there is built-in support for over 100 languages in Plone; basically this means that some people over time translated the Plone interface, so that when you enable the new language, all the "built-in" text that displays in Plone is already translated for you. For example, when you visit a site that is powered by a CMS, you see many familiar features, such as a place to log in, different menu items to click on, and so on. So the great thing about a "pre-translated interface" is that all of the content management side of things has already been translated, so again, you can concentrate on the content. In other words, if you want to add content in ten (or 100) different languages, you wouldn't need to translate "log in" and every other standard element on your website – the only thing you'd need to translate is the titles and text of your articles.

So using LinguaPlone, I immediately set out to enable my ten target languages, and created a simple article in English on how to make a peanut butter sandwich. And then,

for testing purposes, I used the free machine translation service at babelfish.altavista.com, and made translations of the very short article.

Translation Management

The other aspect to Lingua Plone that is helpful for working with translated content is the fact that it has "translation management" built-in. This means that it has a few tools that make it easier to work with making the various language versions of a particular article.

When you enable a new language in Lingua Plone, basically what it enables you to do is to go into a particular article, and "translate into" a new article. It doesn't mean that it translates it for you, but provides you the tools to manage the translated content, right in the interface.

For example, with my peanut butter article, the Lingua Plone interface opened up the English version of the article, and I selected the option to translate the article into Spanish. And the interface opened up a series of windows so that I could manage the English and Spanish versions side by side. In my case, I copied and pasted the English text into the machine translation website, and pasted the Spanish version back into the interface. But the value of this system is that it would allow you to have a professional or volunteer translator log directly into the system and access the content.

Note: In some cases, you may find that you want to keep the original language version of an article in word processing format for translation. It is typical to create multiple column tables, so that an individual row might contain an English paragraph in one column, and the space for the Spanish equivalent in another column, for example. This can also be done in spreadsheets such as Microsoft Excel or the free Open Office spreadsheet (www.openoffice.org). The purpose of using such a document format is simply to make it easier for the translator, who can open the document up, and translate your content on a paragraph by paragraph basis. In the case of a CMS, if a translator preferred this format, you or a colleague could always take the content and "bring it in" to the CMS.

Language Selection – per article

Another feature of multilingual websites that I had wanted to explore (and exploit) was a way to select between different language versions of the site. There are a variety of approaches, including displaying a written form of each native language as menu items to choose from (ex: www.irna.com), or little flag icons. While flags can have political repercussions in some cases (ex: Taiwanese may not appreciate necessarily having to click on the People's Republic of China flag to access the Chinese version of a site.), I felt like it was a reasonable option – and I was in a situation where ANY option was great.

When you have taken an article, enabled the appropriate language, and then translated that article, what Lingua Plone does is to make an additional language flag available at each article.

Language Selection (Global)

I was thrilled to have the ability to work with content in different languages period, and it was nice to have language selection on a per article basis, but what I really wanted is to have "global language selection", so that when a person visited the site, they could pick their language at the top of the site, and then see everything in that language.

Fortunately, someone in the Plone community had developed a tutorial on how to "customize" Plone to add in this ability, to get global language selection. (For those who look into Plone and are similarly interested, see: http://plone.org/documentation/how-to/languageflaglocation2.1.)

CMS Limitations – The Customization Question

One of the limitations (or advantages, depending on how you see it) of CMSes is that they almost always come with some kind of template. The purpose of a CMS is to provide a framework for the development and delivery of content. So the more you wish to customize the "look and feel" of a website generated with a CMS, the more you have to invest in learning how the CMS works. And in some cases, you may end up needing a developer to help.

If your focus is on the content, and you don't need to conform to a pre-existing visual design, then a CMS can be great. If you have an existing website, and in the end, you need the content handled by the CMS, to conform to this pre-existing visual design, then it will take more time and resources – but it can be done.

For me, it is likely that I will not have the resources to do significant customization, so I will be able to make do with the existing "templates" that are available in open source CMSes, but the question of customization is certainly something I'm interested in learning more about.

The Result – How To Make a Peanut Butter Sandwich

So I created the test content at www.ifuedu.com/cftw, and it is still likely that by the time of this presentation, you will be able to go to the url (or an updated one may be linked to from that address), and see the fruits of the exploration.

The result of the exploration is simple, but very satisfying, to know that a content management system can allow an individual or entity to create truly global content, with a reasonably professional look and feel.

Alternatives and the Nature of Open Source

I have been exploring Plone, but heard more recently about some positive developments of a competing CMS called Joomla. It has a complicated genealogy, but in short was an

offshoot of the popular Mambo CMS. At the time of writing (Summer of 2006), a new version 1.5 of Joomla has been announced (www.joomla.com), that hasn't been released yet, but which will address many international and multilingual considerations.

The advantage that Joomla has is that it may be slightly more accessible to more developers, in terms of the underlying code, and easy of customization. It may also be less processor intensive.

The moral of the story is that the open source community moves at Internet speed, and it pays to be out there and learning about material. By the time of the conference, I may be using Joomla to do my research and testing, as compared with Plone. And there are admittedly other options out there that I am not even aware about (please feel free to enlighten me: tkelsey@iit.edu). I have heard that Typo3 is used in Europe and has good support, and I'm sure there are others.

Conclusion

My goal in this article was to introduce the existence of multilingual-capable open source content management systems, and to encourage readers to explore them as a potential way to develop and deliver content to a global audience. At present, there are still limitations faced in having to pay professional translators (or convince volunteer translators) to translate your content, but essentially, the fact that there are open source CMSes period is a great thing, especially when there are ones that are increasingly supporting the creation and delivery of multilingual content. A large corporation can eat multilingual websites for breakfast, using expensive commercial CMSes that one could only dream about accessing. But as an individual, an open source CMS can allow you to go a long way towards creating and delivering content for a global audience, without having to have a big budget.

I wish you the best in your own exploration!

Resources

www.cmsmatrix.org – feature comparison of various CMSes
www.opensourcehost.com – ex: $8.95 USD for a starter account, free CMS installation.
www.opensourcecms.com – allows you to test live CMS installations
www.plone.org – an option to consider for a multilingual-capable CMS
www.joomla.com – another option, especially when version 1.5 is available