Wikipublisher: A Print-on-Demand Wiki

John Rankin
Affinity Limited
Wellington, New Zealand
Email: john.rankin@affinity.co.nz

Craig Anslow, James Noble,
Brenda Chawner, Donald Gordon
Victoria University of Wellington
Wellington, New Zealand

ABSTRACT

Web and print exist as two solitudes: printed web pages often disappoint and converting print documents into good web pages is hard. A wiki makes it easy for authors to create rich web content, but is little help if readers wish to print the results. Wikipublisher lets readers turn wiki pages or page collections into print, with a quality better than most word processing documents. This lowers the time and cost of creating online and print versions of the same content, with no loss of quality in either medium.

Categories and Subject Descriptors

H.4.1 [**Office Automation**]: Desktop publishing; I.7.2 [**Document Preparation**]: Desktop publishing, hypertext/hypermedia, languages and systems, markup languages; I.7.4 [**Electronic Publishing**].

General Terms

Design, Human Factors.

Keywords

Wikis, wiki markup, printing the web, web publishing.

1. INTRODUCTION

Using a wiki makes it easy for authors to collaborate together wherever they may be, as long as they have access to a web browser. This is fine if the result you want is web pages, but what if readers wish to print these? Most people reading more than a page of text will print it; a study of scholarly reading behaviour reports that 80% of researchers read scholarly articles on paper and only 20% online [4].

Reading a printed web page is usually a disappointing experience—we have been conditioned to have low expectations of printing from the Web. Even if the printed result is "good enough" we can still only print one web page at a time, unless the site has deliberately created multi-page articles with a combined "printable view" of the content.

Wikipublisher[1] changes this, by allowing users to turn individual pages or page collections into a document suitable for printing; for example, this paper is also a *wiki page*[2] published under a CC BY SA *licence.*[3] Wiki content is first transformed into XML and then into LaTeX [2], to produce printed output of the highest quality—superior to anything that can be achieved over the Web using CSS or with most word processors.

2. THE WIKIPUBLISHER PROJECT

The Wikipublisher project was conceived in 2004, and the first beta version of the software was released in late 2005. All the software is free and open source. We adopted a number of design principles for the project [5]:

Online First Most of our authoring tools are "print first" and converting print documents into HTML for the Web is hard to do well. Creating content online first makes it instantly and widely accessible without print to web conversion issues.

Print Still Matters The longer and richer the content, the more likely the reader is to print it. Therefore, a web page worth reading is worth printing.

One Authoritative Source Most publishing systems require three or more versions: word processing source; a PDF snapshot of the word processing source, and a set of static web pages generated from the source. The more frequently the content changes and the more authors involved in creating it, the more important it is to have one authoritative source.

Figure 1, *Architecture*, shows the architecture of Wikipublisher. The core architectural decision was to treat generating web pages and generating print pages as separate services. This means one print server can support many web page servers—printing is in most cases a low volume activity compared to browsing, so it is inappropriate to burden the web page server with print duties. We define a print API that lets a web server expose its content in a way that the print server can process. As a result, the print server can work with any web content management system able to support the print API. This design also promotes a more rigorous separation of the underlying content from its presentation in different media, making a wiki an ideal lightweight content server.

---

[1] `www.wikipublisher.org`
[2] `www.wikipublisher.org/wiki/WikiSym/Wikipublisher ProjectPoster`
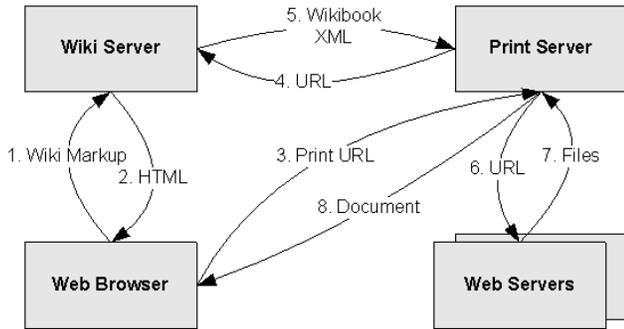[3] `creativecommons.org/licenses/by-sa/3.0/nz`

Figure 1. Wikipublisher Architecture

Authors interact with the wiki server with a web browser (1 and 2). To create a print document, a reader submits a form (3) to the print server which says, "If you issue this http request (4), you will receive a stream of Wikibook XML (5); convert it to LATEX and PDF, then give me back the result (8)." The wiki administrator has configured the wiki server so that, "If you receive an http request (4) in this format, convert wiki to XML (5) instead of HTML." The wiki server thus needs to give the reader a form (2) in order to, "Tell the print server (at this address) to issue this http request (3)." Finally, the print server needs to retrieve supplementary materials, such as image files, referenced in the XML (6 and 7), and return a print document (8).

Wikipublisher translates wiki markup into an intermediate print-oriented XML form, and then transforms the XML into LATEX. We wrote a plug-in for PmWiki [6] (written in PHP) that replaces all the wiki to HTML translation rules with wiki to XML rules. The PmWiki project [3] is a *markup agnostic* wiki engine (almost), which lets a site administrator redefine or augment the markup translation rules. We used the **t**book system [1] to convert the XML documents into LATEX using XSLT. We found that the wiki markup had rules for which there were no equivalents in the **t**book DTD and hence no XML to LATEX translations. We therefore added a range of extensions to the **t**book DTD, style files and XSLT, and called the resulting XML to LATEX conversion service Wikibook and *Wikibook* DTD.[4] Wikipublisher also provides a "print metadata manager" which lets authors and readers customise the way the print output is presented, by passing configuration parameters to the Wikibook PDF server.

We made XML generation and Wikibook transformation as robust as possible. Consistent presentation of printed outputs is completely automatic—not just within a document type (all reports have the same look), but different document types are all recognisably part of the same family. Businesses which typically produce a large number of documents of a small number of document types can get a consistent look (a house style) at minimal cost and in particular with less quality control effort. There is a huge quality advantage when we shift typesetting from the desk-top to the server, because we eliminate local stylistic variations. Of course, limiting local customisation can also be a disadvantage in many situations.

When converting web pages to print, users expect the typesetting engine to apply standard layout conventions for printed material. For a given input, Wikipublisher optimises the quality of the printed output and applies the rules of LATEX typesetting consistently to every page. Authors can focus on content, rather than presentation, and do not need to be typesetting experts to produce professional-looking print documents from their web page collections.

We run a free public Wikibook PDF server for those wishing to try out the software. In the past 4 years we have had 340 wiki sites registered to use the Wikipublisher system via the public server. This has been a fruitful source of feedback for the system's evolution, in response to others' experiences. The web site has an issues register for people to log bugs or change requests, a tip of the week where we publish short "how to" stories, a discussion group, software release notes, and a cookbook for user-contributed local customisations (plug-ins) to extend Wikipublisher's capabilities.

3. CONCLUSIONS

The better Wikipublisher does its job, the less people notice it; good typography is invisible, letting the reader focus on reading. In producing print documents, most people are accustomed to making a trade-off between the convenience of a word processor and the quality of a desk-top publishing system. Most choose convenience, with the unfortunate result that typographic mediocrity has become entrenched in our culture. A big reason for the popularity of wikis is their convenience. Wikipublisher lets us combine the convenience of a wiki with the typesetting quality of the finest desk-top publishing software. Because the system embeds good typesetting practices in the software, the quality is free.

In the future we plan to deploy the Wikibook PDF server in a Microsoft Windows environment; currently it is running in GNU/Linux and Mac OS X environments. For wider adoption, we would like to write Wikipublisher plug-ins for other content engines such as MediaWiki and Twiki. Having the ability to use different LATEX classes would give more flexibility than the four distinct LATEX document types currently supported: letter (with envelope), article, report and book. Finally, we would like to conduct empirical studies of how people are using Wikipublisher.

4. REFERENCES

[1] T. Bronger, 2003. *The **t**book system for XML Authoring*, `tbookdtd.sourceforge.net`, accessed on 27 March 2009.

[2] L. Lamport, 1994. *LATEX: A Document Preparation System (Second Edition).* Boston: Addison-Wesley.

[3] P. Michaud, 2002. *PmWiki: a wiki-based system for collaborative creation and maintenance of websites*, `www.pmwiki.org`, accessed on 27 March 2009.

[4] D. Nicholas, P. Huntington, H.R. Jamali, I. Rowlands, T. Dobrowolski, and C. Tenopir. *Viewing and reading behaviour in a vitual environment: The full-text download and what can be read into it.* Aslib Proc. 60(3):185–198, 2008.

[5] J. Rankin. *Wikipublisher: A Web-based System to Make Online and Print Versions of the Same Content*, TUGboat 29(2): 264–269, 2008.

[6] J. Rankin, 2009. *PmWiki Plug-in: PublishPDF*, `www.pmwiki.org/wiki/Cookbook/PublishPDF`, accessed on 27 March 2009.

---

[4] `www.wikipublisher.org/dtd/wikibook.dtd`