Gentlemen Prefer PDFs

By tim on May 17, 2006

One of the interesting outcomes of our Rough Cuts early access program is some great data on the strong preference of our customers for downloadable PDFs over print books. Based on a little less than 3 months of data, we see that of the customers who've bought Rough Cuts, 60% chose the PDF-only option; 36% chose the bundle of PDF plus print book, and only 4% chose to pre-order the print book only.

These numbers are a perfect match with those reported by the Pragmatic Programmers. Dave Thomas told me in email (and gave permission to post) that in the first quarter, more than 60% of their direct sales were PDFs rather than print books, another 35% was a bundle of print and PDF, and 5% were print only. In other words, when given the choice, 95% of customers want the PDF! This information is also consistent with anecdotal data from alpha geeks like our own Marc Hedlund, who wrote recently on an O'Reilly back channel list:

> "I'm basically up to my nostrils in code for my new company, and I've noticed a very significant change in the way I'm learning about new technologies. There are three tools I'm using almost exclusively to learn (listed in the order I try them): (1) PDF copies of books on the topics I'm learning; (2) mailing lists, and their archives, on those topics; and (3) source code, either from a source repository or Google. (A distant fourth would be project wikis covering the same topics -- the wikis I see are a total mess, out of date, and often seemingly were never right in the first place.) With the exception of system administration topics, where step-by-step instructions matter more to me than real understanding, and Head First/Head Rush books, I'm not using print books at all. PDFs give me the following benefits which print books lack:
>
> - They are searchable. I don't have to rely on the index put together by the publisher -- and that's good, because when I do fall back to the index, it's not useful to me, no matter who publishes the book. Searching a PDF is a huge speed-up over finding something by TOC or index.
> - They are portable. I check all of my PDFs into source control, and without even trying I have them on all the machines where I develop, whether those machines are online at the moment or not. I don't carry anything to or from work anymore -- if it isn't in svn, I don't need it.
> - They are more timely, and often, I can get them in the same hour I find out about them. If the publisher is revising the PDF, either

through a beta program or through a new release, I can often get a new copy of the book very quickly and sometimes for free. With downloads, I can get a cheaper copy of the book immediately, rather than paying Amazon a bunch for overnight delivery of a more expensive print book."

The message is loud and clear. There are a number of important implications, however:

1. As readers choose PDF only, the marginal cost of printed books goes up, as costs increase sharply at low volumes. As a result, there will be a strong impetus for many mid-list books to be made available *only* as PDFs, because only top sellers will have the volume to justify printing. Either that, or print book prices will go up to make up for the increased costs.

2. The results vary somewhat by the *type* of book. See my previous posting on What Job Does a Book Do? Consistent with what I wrote there, reference-oriented books have the highest percentage of PDF-only, and those that provide "fun" are still bought somewhat more in print. Ajax Design Patterns sold 67% PDF-only, while Flickr Hacks sold only 45% in PDF-only format.

There is definitely a publishing opportunity here, and we're trying to figure out how best to seize it. The old gardening advice to "grasp the nettle firmly" seems apposite. We will definitely be offering PDF downloads soon, and would love to hear from you about what features are most important to you, and whether you would tend to buy the PDF only, or whether you'd buy a book-PDF bundle.

P.S. Obviously, our Safari Books Online service offers many of the same benefits that Marc touts -- searchability and immediate access in particular -- but he's right that it doesn't offer disconnected operation. I should point out that it *does* offer an additional benefit that standalone PDFs don't offer, namely the ability to search books you don't already own.

I'll also point out that one of the things that's missing for PDFs is a well-developed distribution system. As I wrote back in 1995 in an essay entitled Publishing Models for Internet Commerce, and in 2000 in a talk entitled The Ecology of EBook Publishing, "distribution systems exist for the same reason that we have alveoli in our lungs. They create surface area.... there are two classes of customers. There are the people who already know that they want your product, who can come to you directly, and then there [are] the people who are going to encounter your product by chance." Web search engines, plus specialized book search engines like Google Book Search and Amazon and maybe even iTunes will eventually offer that kind of serendipitous discovery for eBooks that bookstores provide today, but until there is a rich distribution ecology, downloadable eBooks of any flavor will not reach their full potential.

Part of what we're building with Safari is a channel. We have resellers in libraries, universities, and corporate settings, as well as a strong base of direct subscribers. And that channel is becoming more and more significant. In the most recent 12 months, sales of O'Reilly books through Safari exceeded sales from Borders, making it our #3 reseller behind Amazon and Barnes & Noble, with about 5 times the revenue of our direct sales from oreilly.com. Assuming that we had the same results as Pragmatic, adding PDFs to oreilly.com would more than double our direct sales, but that would still leave it far

behind the level of sales that we get through the distribution channel that we've built with Safari.

When we do offer PDFs, we will probably offer them both direct and through Safari. After all, the essence of my thinking on distribution is that more is better. (We've long made the same argument to bookstores about direct marketing by publishers. Lack of awareness is the biggest problem any book faces.) 51% of our Rough Cuts sales were to existing Safari subscribers, meaning that 49% came direct. However, I would love your thoughts on whether you'd prefer your PDFs through oreilly.com or Safari.