

BookLiberator Prospectus

Brewster,

In the numbers below, I've tried to be conservative on the investment side (i.e., high) and likewise on the results side (i.e., low). If things still look doable even in the conservative case, we're probably in good shape.

The following chart shows our cost-per-page assuming an investment of \$80,000 [1]. It is projected by *[number of devices out there] x [books scanned per device]*, over the course of a year, assuming an average book length of 200 pages (which is quite low).

Our cost-per-page, at varying levels of adoption

Green = below 10 cents per page

↓Devices sold↓ ↓ in first year ↓	←Avg. # of books scanned per year per customer→				
	5	10	20	50	100
-----	-----	-----	-----	-----	-----
500	\$0.160	\$0.080	\$0.040	\$0.016	\$0.008
1000	\$0.080	\$0.040	\$0.020	\$0.008	\$0.004
2000	\$0.032	\$0.016	\$0.008	\$0.003	\$0.002
5000	\$0.016	\$0.008	\$0.004	\$0.002	\$0.001

As you can see, in all but the most pessimistic case (upper left corner) the IA investment brings in results below 10 cents per page -- well below, in most cases. My feeling is that the middle regions of the chart represent the most likely outcome: somewhere between 3 and 2 cents per page.

How likely is it that a customer will do more than 5 books a year? This is all educated guessing, of course, but my own experience scanning with the most recent BookLiberator prototype and two quick-focus cameras was extremely pleasant: I averaged 11 pages per minute without even rushing, and afterwards found that what I most wanted to do was put another book in the cradle and scan again. It's hard to describe how physically satisfying the experience is; I'll bring a prototype out to the Archive in January so you can see for yourself.

One caveat:

Upload time, as you predicted, is going to be an issue. Over a home Internet connection, uploading the page images for a 238 page book took more than seven hours. That was with the cameras set to 7 megapixels, so about 2.7 MB per image. Even with cameras that will go down to the minimum needed resolution (probably about 5 megapixels?), and using greyscale where color is not an issue, the upload times would still be greater than what browser-based uploaders can reliably handle.

The solution is client-side software, but *not* of the kind of complexity we would have faced if doing the entire OCR user interface on the client side. We just need to offer a portable and reliable uploader, offering backgrounding and resumability. This is not difficult; it just has to converse with the upstream server to determine how much of a given upload remains to be done. The web interface would still be the starting point: our web pages would tell the user how to run the uploader, and give out a unique ID for each upload and OCR run (so that those who wish to preserve their anonymity can do so -- we can avoid asking for an email address or even an IP address). The user would pass the ID along to the client-side uploader, which would remember it until the book is completed and then deliberately forget it.

I'll be out in the Bay Area from January 19-26, and will talk to June to arrange a time to come by. In the meantime, if you'd like me to re-make the chart with different input assumptions, I can easily do so -- just send me the numbers to use.

Happy New Year,
-Karl

[1] That's approximately a year of half-time work for me. If it sounds high, remember that both sides of Social Security come of that, medical insurance, etc. What remains works out to about half of my current net income.