LibraryThing:
An Interview with Tim Spalding

LibraryThing is an online application that enables users to catalogue, tag, search, and sort their books, as well as making book recommendations based on the collective intelligence of other users' libraries. The site searches all five national Amazon sites and over 80 libraries worldwide that provide open access to their collections with the Z39.50 protocol (used in bibliographic software, such as EndNote). LibraryThing is also a blossoming social network, connecting people with similar libraries and interests. Since launching in 2005, LibraryThing has accumulated over 619,000 users, with 36-million books catalogued and 47-million tags applied.

Crossroads caught up with Tim Spalding, the founder and lead developer of LibraryThing, to find out more about the social networking site for bibliophiles.

Your background is in the Classics. How did you go from studying Greek and Latin to being a Web developer and Web publisher? Actually, it's something of the reverse. I was a computer kid—really the first generation of computer kids. My dad got an Apple II in 1979 or 1980 when I was eight. So I was programming very young. I focused on text adventures, so in a way I was also atypical—computers back then were for math. When I went to college, I decided to let computers to the side for a while. Georgetown didn't have a good program, and my academic interests were otherwise. Majoring in computers was for math-heads who didn't read. I had no desire to take courses on Assembly Language or even C. I worked in Basic and later Pascal.

In college and after, I did do some computer work, but it was mostly in design and other applications. After the computers stopped shipping with languages—which happened early on and now happens again, as every Mac has Perl, PHP, and Python at least installed—I stopped writing in them. But I did the Filemaker database for an archaeology dig I was on. And, at graduate school, I even wrote a program to parse Latin meter—still quite great, if you ask me—in, get this, Filemaker.

When I left graduate school, I took a job at Houghton Mifflin "managing projects" in their instructional technology division. I had enough background to realize that the projects were in trouble because they were ill-managed, but because nobody knew anything about computers. For example, there was this whole process involving manual retyping of handwritten information on a PC, because the PC program couldn't read Mac files. I realized that it was a line-break issue, so I got back into programming to futz with text files—choosing Perl, of course. I spent the next 2+ years there learning everything I could about everything. I rose rather rapidly and was soon running a sort of start-up within the department, with four employees reporting to me about eBooks. Abby, LibraryThing's first hire, was first hired there.

I moved to Portland to work on my fun but labor intensive Isidore-of-Seville.com site. Basically, the idea is to make authoritative, highly comprehensive vertical sites on topics, and make a buck or two per day off of Google ads. It worked fairly well, but the sites proved harder to make than I'd like. I stumbled into making LibraryThing as a side project. I did think it would make money, but I was thinking on the order of $10/day, and I figured I'd do it and finish it.

It's hard to imagine LibraryThing as a pet project. What was the point at which it exploded into a business venture? How comfortable is the pile of gold upon which you now recline? As I made it, I realized it might turn into something modestly successful. It was so much better than anything else. Still, I was focused on the cataloging aspect and didn't see it as a social project until a few weeks after launch. Indeed, I tried to partner with another site that did some social interaction but had crap cataloging (Bibliophil.org, which is still around). After a few weeks, it hit the Guardian online, and from that point, I was sleeping with my laptop.

It launched in August 2005. After a few months, I wanted to get some financing, eventually coming to terms with Abebooks. It was a small investment, but the terms were generous, and they were interested in growing it.

How comfortable is the pile? Well, on paper, I've got a lot of money, by my standards anyway. But I have to sell the company to get it. I don't want to do that anytime soon. I'm not paid so well. When [Cambridge Information Group] bought a small percent of the company, I sold a few percent. That's given my family a down-payment on a house, which was very good to get.

LibraryThing has rapidly grown into a large project with a huge repository of book data, drawing a lot of traffic from its many users. This must take substantial resources to support. Tell us about the technology behind it all. How much software is off-the-shelf, and how much is in-house? Well, it's a fairly standard LAMP setup—Linux, Apache, MySQL, PHP, and Python. We also use memcached a fair amount, and other open-source solutions sites of our sort have (e.g., Nagios for monitoring). Apart from some laptop software—Quickbooks, BBEdit, Excel—we don't have any for-pay software. I don't know what distributions and such. We have a sysadmin, John, who does.

Scaling is the great problem of a site like LibraryThing—both financially and as software. We've moved through various scaling and caching issues. I have a really good handle on MySQL issues—what to normalize and what not to, etc.

As far as boxes, we have something of a muddle. The other big problem with organic growth is that your needs change, but you can't

afford to throw away old stuff and standardize. We are about to basically double our resources. The new boxes are much more thought-out.

LibraryThing now has nine employees, including yourself, whose job titles range through Librarian, Library Developer, Developer, and SysAdmin. I would guess you have people with a mix of backgrounds. Do any difficulties arise from this? How knowledgeable are the librarians about the technical side of things and the developers about the library side of things? Interesting question. (Thanks for asking ones I don't usually get.) Mostly, it's not a problem. We have two people who do mostly library-related development, for our LibraryThing for Libraries product or back-end work on getting and parsing library data for LibraryThing.com. Chris Catalfo has an MLS [Master of Library Science]. Casey Durfee does not, but worked for a number of years at both SirsiDynix and the Seattle Public Library, so he gets library culture.

Our other two librarians, Sonya and Abby, do not do much "technical" work for us. Abby's job, as originally conceived was "everything but the code." By the time I hired her, I already knew more about library data and perhaps even cataloging than she did. But she was helpful introducing me to the library world. She is, more generally, an exceptional person—smart, hard-working, etc. Library school may have taught her some of that, but she got two master's, and I suspect the History master's was at least as informative.

Notably, although Abby doesn't do much technical, she does edit HTML, and she has a very good understanding of MySQL, useful for various account issues. At one point, I sat down to teach her MySQL and, within three hours, I had gotten her to the SQL query for "people who like this book also like those books." Since then, I've often found prospective developers can't get there.

Sonya was hired specifically to talk to (well, "sell" to) libraries. We needed someone personable who understood library culture. Notably, she's married to a Ruby developer, so she gets the other side too.

Our other developers have a more mixed understanding of the library side. Christopher Holland, an early hire, has an intellectual/academic bent—he did computer stuff for archaeological digs too. He may not know much about the library side, but he understands the content. Mike and Luke have a more casual relationship to libraries.

One thing is clear. I've hired a number of non-library developer people—developers with no knowledge of the library world. If they work on a library-related project, I need to introduce them to the library tech "world," such as it is. They are universally stunned by how backwards it is. To anyone who follows the fast-moving, open world of contemporary startups and open-source projects, the library tech world is a cringe-inducing backwater. The technology is Jurassic, and so many of the "tech leaders" simply don't get it. Within libraryland, for example, there is debate over whether libraries should expose their catalogs to search engines. I recently got into a fight with one [Online Computer Library Center] employee, a "tech leader" of great renown, who called the idea "lunacy." In the rest of the tech world, that opinion would be laughed out of the room. There are any number of similar examples of libraries not understanding technology, and the Web in particular.

The other side is also clear. Many non-library people don't understand just how much great data libraries have, and for how many people libraries are still central to their lives. LibraryThing's mission is to give libraries their due—to show what they've got and mix it up with innovations like tagging, etc.

What's a typical day (or week) for you? My time is split between developing and "business"—managing employees, answering stuff like this, etc. I do a lot of talks, but they are mostly bunched together. I do some conferences.

LibraryThing is closed-source, and details of your recommendation algorithms are not made public. How much do you think other domains (e.g., movie-, recipe-, photo-sharing) might benefit from learning what goes on under LibraryThing's hood, and how much is specific to books? We've got a few algorithms that might help others improve, notably our "special sauce" and our tag-based recommendations. But the main recommendations—and I suspect [the others too]—do not differ much from standard recommendation algorithms. If LibraryThing works, it's because our data is so much better—it's really you, not some [stuff] you bought or some movies you rated.

Books differ in a few ways. First, there are just *more* books—way more books than there are movies or CDs. Second, you experience books differently than you do movies. I see a lot of movies I don't like, but I mostly buy books I like. "Chick lit" is for chicks, but a "chick flick" is for guys to go to with their girlfriends, to suggest they're sensitive or to make up for making them sit through some action film.

There are a couple of consequences of the differences. People often wonder why I don't use ratings (much) in our recommendations. The answer is simple: by and large, people buy the books they like. If they don't like them, there's still a reason they bought them—they thought they'd like them. In a way, people own the books they "deserve." Of course, you might have some gift book you hate but won't get rid of. In reality, those are a small part of anyone's library. In fact, most people rate books in the middle. The average is 3.8, and the standard deviation is low. We do a lot better looking at ownership patterns than using ratings.

One factor people never get is that LibraryThing isn't trying to sell you [stuff]. Amazon is, and it results in some noise. For our purposes too, we're trying to entertain you most of all. So, for example, it may be statistically true that if you like [Harry Potter 1], you'll like [Harry Potter 2], but showing that—and 3, 4, 5, 6, 7—is boring. People already know it. So recommendations are as much about being interesting as being predictive.

How many of your algorithms are invented in-house, on-the-fly, and how much have you been able to use existing techniques? Speed is always hard with recommendation algorithms. I wrote them all myself, from first principles. I've looked at some formal algorithms for doing it, and find they are more elegant in some ways, but also more limited. I suspect we'd benefit from some help there. I've been meaning to take a statistics course, in particular to handle tag-recommendations better.

Bug reporting and suggesting site improvements are lively topics on the talk forums. Who decides what fixes and changes are made to the site, and how are they prioritized? Anything big, I have a hand in prioritizing, and we have periodic meetings to lay out what people are doing—where everything is debated, but, ultimately, I make the final call. But all the developers have a hand in listening to bug reports about fea-

tures they developed. I try to give, and developers take, a lot of freedom there. Chris, in particular, has been known to work for a week on some new feature or bug fix which I didn't ask him to work on, and, indeed, for which he's ditched what I gave him. Mostly he does good there, but it can be annoying.

What new and exciting features can we expect in the near (and not-so-near) future? That would be telling.

We'll get Facebook in a month or two. We have lots of fun plans. I hope that we have the new widgets out by the time you read this.

LibraryThing looks like a good flagship for collaboration, data sharing, and their mutual benefits: libraries share their data, and you offer them recommendations and social data, e.g., tags. Your social data would also be interesting to people researching social networks, for instance. Do you share data with non-libraries? We have a standard form for students studying social networks and etc. We'll give out most data under those terms. (The [Terms of Service] even mentions we may do that.)

If you look at the **API page**, you'll see we share a lot of our data, most of it for free. We'll be doing more there soon.

What about LibraryThing are you most proud of? What would you say is the most important contribution of LibraryThing today? I'm very proud that, basically, I invented book-based social networking and social cataloging. I proved that it was a good idea, and that people would be interested. (Some early critics suggested there were only a few hundred people who'd want to it.) I'm very proud that we've always pushed the envelope. Common Knowledge, our Talk system etc. Tagging in general makes me happy. We've done so many groundbreaking things there—we come up again and again in the recent book on the subject—tagmashes, tag combinations, tag-subject mixing, etc. LibraryThing proved that tagging would work for books—work like Hell, I might add.

I'm also proud that LibraryThing swims in the opposite direction on the basic issue of intellect, respect, and integrity. We don't require e-mail for the privacy-sensitive. We don't show little square photos for user names—we use *words*. We take it for granted that people want to edit their data, and to get it out again. We respect libraries. We enforce a no-ad hominem rule. We run the site collaboratively as much as possible. We have no "community managers" (shiver). We talk openly about money. We don't have any advertising for members. We charge.

Much of this runs counter to contemporary wisdom. Shelfari and Goodreads don't allow you to edit your book data, and they only use Amazon. (If you're over 30, you have books Amazon doesn't sell!) They are less intellectual, more pushy in their virality—Shelfari famously so. They were built to flip.

We've actually been hurt a lot by them. Stats are very wiggly, but Goodreads is probably larger now—in sane metrics, not bogus ones. But we're still growing quite nicely, and getting deeper and more interesting. If, in the future, the market settles down into the free one that treats you like a child and the paid one that treats you with respect, I'll be fine with that.

What is your vision of the future for LibraryThing? I want every serious book person to use LibraryThing—every day, every week,

whatever. I want it to serve as an essential resource for information and socialization for everyone in that peculiar demi-monde, linking up every major player and party—readers, authors, publishers, agents, new bookstores, used bookstores, book clubs, public libraries, academic libraries, athenaea, literary societies, etc.

Finally, and most importantly, what's on *your* bookshelf? Ha. I'm listening to Philbrick's Mayflower [2] on this trip. At home, my nightstand has Problem Solving 101 [3], What Have you Changed Your Mind About [1], and two books on ancient history.

—Anna Ritchie

References

1. Brockman, J. 2009. *What Have You Changed Your Mind About?: Today's Leading Minds Rethink Everything.* **http://www.librarything.com/work/6839567**.

2. Philbrick, N. 2007. *Mayflower.* **http://www.librarything.com/work/806107**.

3. Watanabe, K. 2009. *Problem Solving 101: A Simple Book for Smart People.* **http://www.librarything.com/work/8072335**.

Biography

Anna Ritchie will graduate with her PhD in Information Retrieval from the University of Cambridge in May 2009, having previously received her Master of Philosophy in Computer Speech, Text and Internet Technology and BA (Hons.) in Computer Science there. She now works full-time on stress testing her bookshelves.