



Chris Palma on Google Book Search at the Text & Academic Authors Association Annual Meeting 2008

M: I have the pleasure of introducing an old friend. I met him last year, and when you meet people at these conferences, they're old friends immediately. Chris Palma, who you can see so you don't need me to tell you, is the strategic partner/development manager of Google Book Search, and I'm going to make all of my introductions today err on the side of brevity because I think the best way of showing respect to speakers is not to talk a lot about them but just to let them talk.

But I am just going to take one moment for one very egocentric part of this introduction to say that I had never heard of Google Book Search last year and I was absolutely fascinated by the presentation, and I immediately took charge of my life and sent my books to Google and waited for them to do the old-fashioned kind of scanning, a page at a time, with their machines, I guess, to get it in, figuring it's better to ask forgiveness than permission, so I didn't want to ask my publisher to do it, because what if they said no? And then we discovered later that the publisher, well, of course was going to do this and did it redundantly and was told the author already did it.

And the really wonderful part of the egocentric story is when I go on their website and I check things, first of all, I see the many, many people who are looking, but also, more importantly to me, they make my book look so much better than it really is because the selling point of my book is it's more visual than anything in the competition and every single one of the sample pages that they have to show you about my book is one of the most visual moments from the book.

So, they wrote the book that I should have written and so what can I do more? I look forward to doing wonderful things for everybody else like you did for me. Thank you so much.

PALMA: Thank you, Paul. Thanks very much.

(applause)

PALMA: I'm looking out here. I'm seeing a few familiar faces from last year. How many people were here last year? At my session?

M: In Buffalo.

PALMA: In Buffalo. So you folks really have a bipolar thing going. You're going Buffalo and now you're doing Vegas. I for one, and my wife, are very happy that you hosted this here this year because for once I dragged her along to a work-related event.

I was asked by Paul to try to reconstruct to some extent the tenor of last year's talk and take it a step further. Most of you were there, so most of you have a better sense of what Book Search is. We covered the 1.0. We covered what Google is and what our business model is. But I wanted to take it to a 2.0, so this is going to be Google Book Search: The Empire Strikes Back.

It's going to be going beyond Book Search, and I'll review just a few basic tenets of where the program is today. We've made some substantial developments, some interesting developments in the program since I talked to you last, and I'll review those as well.

And then we'll go further and we'll talk about other Google-related tools that I'm not in charge of at all, being their not part of my product team, but I know a little bit about, that are very useful tools for publishers and authors to try to profile your books in this Web-based ecosystem, and talk a little bit about the interconnectivities between certain products like Book Search and Blogger and YouTube and all kinds of other things that are emerging on the Web and driving traffic.

Last year I spent probably half the talk really introducing you to Google and what we do, and I'm not going to do that here other than to say, this is our view of the world. We've got 1.4 billion people online today, and that's growing at an enormous clip.

I was in New Delhi about a month and a half ago trying to bring those publishers the Book Search program. It's one of the fastest growing Internet audiences in the world, as you might imagine, but some interesting things are going to happen in India in the next three years. A mere 1.2 billion people, and publishers said, gee, there's only 110, 115 million people online. That's all there is, so how can you help me?

I said, well, that's actually a pretty sizable number still, and it is the third largest English-speaking country in the world. But they're at close to 90% saturation for cell phone usage in India, so they've leapfrogged over certain things like land lines and other things like high speed.

We're building mobile platforms right now for Google. It's an open platform. You might have read about it.

When cell phone companies begin to make open Web platforms part of cell phones, you're going to see the adoption of the Web explode in places like India and the underdeveloped world. So now my pitch to publishers is, imagine the exposure and the democratization of knowledge that's going to occur when we take that next leap. So that 1.4 billion is really important to us, but it's growing all the time.

And we look at commerce all the time. Commerce is a huge part of what people do on the Web and obviously it's a huge part of how we make our money. But sort of an interesting number that you would think I'd be spouting five, 10 years ago is that last number. 183 billion people doing e-mail a day. Two million e-mails every second.

What does that tell you about the Web? That the Web is enormously social. The social aspects of the Web are going to be crucial as we start to explore Web 2.0 and how to leverage the social aspects of the Web to get our ideas out there in the world.

So a quick review, or not.

Google Book Search today is 1.-something million titles from partners. We have over 110 million publishing partners worldwide and many, many authors on top of that are in the program.

We source books, as you know, from two completely separate projects, one being the Library Project. We partner with academic and public libraries all over the world. There are something like 25 now major libraries, like the University of Michigan and Harvard and a few in Europe, and that's where we go in and scan books in the library, some of which are very, very old, some of which are in the public domain, many are out of copyright.

But what I deal with mostly is the Partner Program, which is authors and publishers who work directly with Google to find ways to surface their books in the program.

Top-line facts. The important thing to mention here is that we're expanding internationally at a very rapid clip. Over 51% of our traffic in Book Search is coming from overseas now, and so we're seeing rapid adoption and growth in traffic coming from places like South America where book distribution has traditionally been quite poor and book marketing even worse. Places like India and China and the Pacific Rim are growing at a rapid clip, and we're just working to catch up.

We're in about 70 international domains, maybe about 80 now international domains, and we're also working to localize in local languages the entire Book Search interface. The idea is that we want to get to in the next five years 50 international languages. We're probably around 40 right now. So literally, localizing the language of the interface for those markets as well.

We have books pretty much scanned from every language you can think of under the sun. There are still two or three that stump us, like Hindi and Urdu and a couple more, but we're working to actually upgrade our scanning capabilities and technology to get those, and it's very important that we speak to a global market.

So real quick, how does Book Search work? Exactly the way Web search works, only instead of indexing the Web, we index and search through pages of books. So every single page of every single book, consider that a Web page. Our index goes through and does a full text index of every word in the million-plus books in the program. Just like Web search, delivers a snippet of text along with the book jacket, to give you a sense of whether you want to click on that book or whether that's the right book that you want to click through on.

There's a relevancy ranking there to those results, which is top-secret stuff that we don't share with people, but there are some obvious ways that people can make sure that their books appear in those. For instance, making sure that they're in the index to start with.

And then once you click through like a Web page, you're brought to our browse pages. Our browse pages are controlled in terms of the browsability factor by the author or by the publisher. The publisher or author determines what percentage of that book actually gets browsed. Our baseline is 20%, because we realized when we showed fewer than 20% of the pages in the book, people tended to get frustrated and actually leave the book.

We've been studying this for three years now and looking at user behavior, and 20% is the bare sweet spot. You have to start with 20% viewability in order to get somebody to actually be interested in the book and drive further interest in the book and hopefully click on a buy-the-book link as well.

There are portions of the book that we block at all times, and that's dynamically generated blocked pages, and anybody that uses Book Search intensively will learn over time that you can actually do a search and land on a page that says, this page is intentionally left blank. The actual query terms are still there behind the page, but we're kind of sending a message to the user that you're not going to ever get to see this, a large part of this book.

And we further message it in what engineers call the interstices, so the stuff between the pages as people are browsing, to say, preview only. So we send very

strong messages to the user that, don't expect a full-text experience, but expect a browse experience. It's very important to set the user expectations.

And then, of course, we disable copy, print, paste, etc. So all the user's really getting is a low-resolution image of the page.

And then the publisher or author can choose to add and subtract books at will. So a publisher loses rights, reverts rights back to the author, they simply notify Google and the book comes out. So it's a pretty low-risk proposition all around.

The whole driving force of the program is to build in new features and functionality to further drive people towards clicking on a buy-this-book link. So much of what you see related to About This Book pages and the excerpt pages themselves are really with the intent in mind to drive the user to click on a buy-the-book link. So the publisher is always the first link and then we list other booksellers, depending on where we might be in the world. And we brand the publisher as well in a couple of different places.

So any questions about basic Book Search at this point? Pretty clear to everybody? Kind of covered this last year? Yeah.

M: I'm curious. My book has a lot of cartoons in it. If there's a cartoon thought bubble, is your scanner treating that as text or does it more (inaudible) that thinks it was graphic?

PALMA: It depends on how you treated it.

M: Could you repeat the question?

PALMA: The question was, let's say a book has half-tones or line art. Maybe there's some text within the art itself of cartoons, thought bubble or something. Do we index that text as well?

If in fact it is represented as text, we do index it. But if in fact it's a half-tone and it looks to the crawlers as if it's an image, we wouldn't actually do that. At least not now.

Yeah. Over here, over here. Sorry.

F: I'm sorry. Could you tell us a little what it means to add your brands on the page? I don't know what that means.

PALMA: Sorry. Simply, the logo at the bottom here. Here's O'Reilly. Publishers are obsessed with their brands. They want to make sure their brands are always appearing. They want to make sure that the colophons look good. For the most part, publishers don't have brands. Authors have their brands. But nevertheless,

for the sake of the publisher, we put a link back to their home page and a link directly to their product pages as well to buy the book.

So the link is here where you see O'Reilly. Where you see O'Reilly here is the link directly to the product page for this book on O'Reilly's site. This here is simply a link back to their home page as well. That's what I meant by branding.

F: Thank you.

PALMA: This gentleman.

M: You mentioned there's a 20% minimum (inaudible). What's the maximum (inaudible)?

PALMA: 100%.

M: Really?

PALMA: Yeah. I've got a slide for this later on, but we've crunched the numbers and there's a direct correlation between the number of pages that are viewable and the propensity for somebody to actually click on a buy-the-book link.

Think of it if you were in a bookstore and you picked a book off the front table and you started to browse through it and you got to a certain point and the clerk came over and said, sorry, that's enough. Buy the book or leave. Everybody's going to leave, and that's exactly what we found.

So publishers are actually doing an about-face here and actually finding ways to increase the browsability of their book. Macmillan, for instance, did a study where they found that where someone looked at fewer than six or seven pages, they rarely clicked on a buy-the-book link. So their magic number was six, so all they thought about now was how can we drive people to look at more pages, and one of the ways to do that is to drive browsability up.

What's nice about this is the publisher has a sandbox to play in here from a marketing perspective. They can go and turn particular books, ratchet them up to 80%, 100%, take maybe (inaudible). Let's go with the 100% model. See what that does. Look at the traffic. Experiment.

Because for the most part, a book that's actually only viewable in a low-resolution image on the Web is not all that useful to people.

Yeah.

F: Do you have any statistics that tell you how many people buy the books versus (inaudible) from your site?

PALMA: It's a very good question and it's the one metric that we actually can't track, because we only track buy-the-book clicks. Once they click through to the publisher's site or to Amazon or to anyone else, we're not privy actually to the conversion because no one shares the conversion data. We don't know if the person actually did buy a book or not.

Publishers like Simon and Schuster who have tracked these things and others have told us that we convert where they can test it on their own website. Traffic coming from Google Book Search converts at 15 times the rate of traffic coming from anywhere else on the Web. And that's a massive number. The average is 1% to 2%.

But it's not a massive number when you consider the fact that somebody was browsing the book and intentionally clicked on a link that said buy this book. So you have a pretty qualified user at that point. The only thing that might turn them away at that point is price.

But I'm sorry. I don't have any particular figures. Any other ones before we move? OK. Good.

So Google Book Search, like Map Search, Product Search, Image Search, all fantastic. Individual indices sort of sitting in Google. Does anyone go to the More tab and find those other indices on Google to search on? Very few people do.

This is where they start. The vast majority of our users start on this tabula rasa and their intentionality grows from here.

The tricky thing about search technology is that it's inductive. It's learned over time. It's not like you can build the perfect algorithm for search and solve everybody's need when they punch a query into Google. You have to actually built a pretty good algorithm, see what people do, see what they query, see what the results are, see where they click, and then tweak those to create more relevancy.

So it's very important for Google to sort of try to figure out what people are looking for and what type of content they're looking for.

F: (inaudible) talking (inaudible).

PALMA: Yeah.

M: Google eavesdrops, and that's what you've got to do (inaudible) and so you can see what people searched for, what came down and what they clicked on. What

you can't see is whether they're clicking in exasperation and running away or if they found what they wanted or not. How do you tap that?

PALMA: You tap that through analytics about what they did once they clicked away. You could actually see did they come back and click again. And this is all nondifferentiated data, right? So we don't know who it is that's doing this. We know that it's done, and from there, we can learn something about what people did.

If you had search results for Khmer Rouge and the first five results were never clicked on and somebody always clicked on the 15th result, why is that? What was unique about that 15th result?

It's tricky business and that's why we have the top software engineers pretty much in the world working on this day and night.

So part of this initiative to figure this all out is what we launched typically and very quietly about September timeframe of last year, something what we call universal search. The idea is that – and I'm sure you've see iterations of this already in your searching.

You're searching for New York Dental School and suddenly you see a map come up. The algorithm intuited that a map would be very useful in this situation. Something about that query triggered a map result blended right within the Web search results.

So you're not seeing Web only. You're seeing actual content, because map content is like book content. Somebody from Google's got to go get it. So it's the idea of intuiting content types as well as relevancy of the query results themselves. Tricky business, and we've started to do that very profitably already.

One of the very subtle ways to do that – I talked about the More tab. Before, you actually had to click on the More tab to find books. If you didn't know you were looking for a book, you'd never see Book Search. But our promise has always been, let's take an intensive user seeking information, very specific information, and if a book is the best relevant result, let's surface a book, because then we're taking someone who never knew they were a book buyer and converted them into a book buyer, because books are some of the most relevant sources of information on the planet. We want to make sure books surface in those results.

So one of the easy ways – well, not easy ways to do it, but logical ways to do it – is to actually force them to a filtered result. So instead of them having to think, oh, maybe there's a book result here. I should go to Book Search. Maybe they never knew about Book Search. We actually trigger, when appropriate, a Books or Images or News result at the top for them to filter down and actually look at books only results.

Then once you click on that result, that filter, you're going to go directly to books.google.com and see books only results. So we're driving an enormous amount of traffic today to books in this fashion.

Another way we're doing it, and this drives an enormous amount of traffic as well. Somebody queries an exact or close-to-exact title of a book. This is extremely valuable to publishers because when we have what the engineers call a strong signal – this is a strong signal for a book – we're actually surfacing something better than the appearance of a Web result. We're actually surfacing a book jacket, so we're giving a stronger indication that there's a book here that's quite relevant. So without either having to go and filter for books, you're actually presented with a book result itself.

The same on the author side. Once you enter the author's name, if the book is in the Book Search index, we'll actually surface what we call a one box, an author one box for all titles that we have available in Book Search.

And then the fanciest science of all is to actually blend book results in with Web results. So here's where we want to say, something on page 125 or chapter 12 of a book actually is quite relevant to that query. It's not an author query anymore. It's not a title query anymore. The query words are not in the metadata anywhere. But inside the book, the algorithm now understands that a piece of that book or even a page of that book answers that query spot on.

It may not get the first result, but it may be one of the results that you see.

Here's an instance. I'm going to go back one. Here's an instance where – and I apologize for the kind of superficial examples here, but it's a nice one to kind of illustrate what we're talking about.

Organizing sock drawer. You've got position one, two, three and four. The first three positions are all websites. The fourth position is how to organize just about everything. So now you're thinking of someone who said, gee, I want to organize my sock drawer, but I'd really like to organize everything. Anybody that organizes their sock drawer must be a maniac, so if they're going to organize their sock drawer, they're going to organize their kitchen, they're going to organize their work bench, they're going to organize everything.

So we've surfaced a book result to serve that need. Now you're bounced right into the book itself and to the page in the book that deals with the search query. And not only getting organizing your sock drawer but sort your sock drawer and organize your shoes.

This is the promise, because we're getting better at this all the time. This requires massive amounts of traffic and massive amounts of users to be doing this in order

for us to get better at it, but 50% now of the traffic coming to Book Search and to books is now coming from some version of universal search. In other words, coming from google.com now, whereas that was a much smaller number even in September.

Any questions about that, because this is really important. Yeah.

F: So this incorporating the Book Search into a regular search, is Google going to do that for other things that are under More, like Google Scholar?

PALMA: Yeah. We already have.

F: You already have?

PALMA: Yeah. All these things are already starting to happen in all the other indices, News Search, Blogger, everything.

And it is tricky stuff, so you'll do it one day and you'll get a certain result and there is the book in position number three, and then you'll do it another day or from another part of the country and it won't be there anymore. Maybe it'll be on page five or maybe it'll be higher. It's something that we continually tweak and are continually learning about as time goes on.

But that's been the promise all along, that Google's going to blend all these indices into one search mechanism.

Yeah.

M: Google dropped a couple of indices recently. Can you talk a little bit about that, because there was a time when that was a little bit of a thorn in the side of authors, the library, I think, indices. There were a couple of projects that Google started to scan books in libraries. From what I understand, those have been dropped. Can you talk a little bit about that?

PALMA: No, that hasn't been dropped at all. Those are still part of – you might be thinking about Microsoft Live Book Search, which closed its doors on their book scanning from libraries.

M: It was an opt-out project rather than an opt-in project that Google had.

PALMA: If it's the Library Project that you're speaking to, maybe.

M: It was the University of Michigan and it was some others that were participating.

PALMA: Yeah. Nothing's changed. We continue apace to scan books from libraries.

The opt-out piece of that is if, for whatever reason, a publisher said, gee, I might have books in those libraries that may still be in print. I want to be able to send a list to Google to opt out of that library program and make sure that my books are not scanned as part of the library program, they can do that. But that's been the case since the beginning.

M: But that project still exists?

PALMA: Oh, yeah, yeah. It's become much more internationally focused because once you have five or six major research libraries in the United States, there's so much overlap in what you're scanning. We're trying now to reach out to other languages, make sure that the French are not mad at us for cultural imperialism. We want to reach out a little bit and get some of those other libraries into the program. But no, we haven't shut those down at all.

Yeah.

M: I'm just interested in the global aspect to this. You mentioned all the various languages, but just breaking down a long (inaudible) of video into, in the text. Are you doing likewise between, say, Google.ca and Google.uk and all the rest? Are we starting to see those results also (inaudible) US (inaudible)?

PALMA: Yeah. To the extent that the publisher has rights, or the author has rights to actually sell the book in the US, we actually have territorial filters, so when the publisher puts their list into Book Search, they can actually prescribe where the book actually surfaces.

And that's been enormously useful when we're pursuing publishers in Malaysia and the Philippines and India who have never, ever had any insight into how to reach the Indian market in the US, for instance. Those books are now surfacing in the index in the US. If that's where you're going with that, that's already happening today.

And by the way, the triggers and the filters and all those things behind the scenes that are going on on the algorithm side differ from territory to territory. There are definitely regional factors that play into search results. So one result in the US, you might not get that same result even where the book is in the index in both places. You might not see it in the UK, or at least see it as high, because there are other regional factors that are playing into the search results.

So we spoke to this a little bit about opening pages up. What we've found is that there is this direct correlation between the number of pages that are visible to the user and the click-through rates.

Now, this fuzziness up here, I originally thought that must have had to do with the fact that after they see so many pages, they don't need to buy the book. But I was

informed by our product team that, in fact, it's only fuzzy because there are so few publishers that allow their books to be viewed in this fashion. So where you have a million, million and a half books in the program, you just have a far smaller number of books that are even viewable up here, so they're not opened as much at the higher end, so you tend to get, from a statistical point of view, you tend to get some fuzziness up there.

But you can see the trend, and it's pretty strong.

We want to talk about what we're doing for publishers. I know this is an authors' organization, but in some ways, I hope that you'll be my emissaries back to your publishers to say, did you know these programs are actually available to you for free and they will help drive traffic to our book and it'll help make your website more valuable?

We offer these tools to publishers. We work enormously hard on the back end to make them godawful simple to implement. The uptake is modest. So I wanted you to know that these things exist.

The first obvious thing is, any publisher with books in the program can implement co-branded search for their own site. By co-branded, I mean from the user's perspective, they're starting their search from the publisher's site and the results are only the publisher's books.

Think of every publisher that's in Book Search for their own website, having a Search Inside the Book capability, a capability that took a certain online retailer years and millions of dollars to develop, and we're giving it away to our partners.

Here we're searching from within the Cambridge site and getting search results, again, that are branded for Cambridge. What you can't see here is that if you looked up at the URL, you're actually looking at books.google.com, but for the most part, users are not going to look at the URL and they're going to still feel like they're on Cambridge site, and we're going to have search results that are filtered to show only Cambridge books.

And then we're going to have a preview experience, again, completely branded for Cambridge University Press.

Now the buy link is Cambridge's buy link. We don't show the other retailers because we've surfaced this result and the search from the Cambridge site. So it's a full cycle back to their product pages to actually buy the book.

We also allow the publisher to put a Search Within This Book button. So let's say that somebody landed on a page deep in the publisher's website. Maybe they landed directly on that page from a Google search. They may not want to do a full text search, but they may want to actually search within the book itself in

order to see if that book answers their query, if it's a book that they really want to buy.

This is kind of a sneak preview because we haven't – I think we've implemented this with one test publisher. But one of the problems with implementing a Search This Book functionality on your product pages is – let's say that you have 3,000 books in print. You have to embed a little strip of HTML code in every one of those books every time you want to implement full text search.

So now we'll have the capability for the publisher to implement what we call an availability API to automatically generate a preview button anytime that the book is both simultaneously on the publisher's product pages and in Book Search.

So let's say that the book was delivered to Google tomorrow and we got it live in Book Search in two weeks. At that two-week time that it went live, anybody running that API, we'd automatically populate this browse button and this preview button on that site.

Stanford University Press is actually one of the test cases here, and they're kind of going both ways. They're implementing the preview button, which is more of a browse feature. You click on that and go to the title page, or they can search within the book itself without much heavy lifting at all.

There's our browse feature.

We're also, by the way, giving this capability to retailers of all sorts. So let's say that Borders launches a new website, because they did about a month ago. They're a little late to the game, to the party, but they did launch their own retail site about a month ago.

Let's say that Borders came to us and said, Borders is hurting a little bit. We don't have the infrastructure and the money, basically, to build this technology, but we've got to compete with Amazon and we'd really like to have a preview function. We're now building a way for them to, again, put this preview button on every book that happens to be in Google Book Search.

You can imagine how many books they have on their site. It's probably 750,000 or something books that are on their site. We've got a million-plus. Chances are, we're going to have most of those books, but it will allow them to actually build just that preview component. Google doesn't show up anywhere other than calling out to those excerpt pages to show right below their product pages. It's just another way to drive more adoption and more sell-through of that book.

But the experience of viewing those pages on Borders' site doesn't change from Google's site, because it's coming from Google site.

Everybody follow that? So if you set the viewability at 30%, somebody seeing that book from the Borders site is still only going to see 30% of the book and then be logged out, right?

So you can see kind of where we're going with this. It's a way to begin to, in essence, syndicate Google Book Search. As much traffic as Google gets, we want to make sure that every other place that somebody was searching for a book or searching for an idea, we can surface a book from Book Search and actually drive a sale.

M: Just for clarity. I don't think it's terribly important, but is that always the same 30%? In other words, the 30% you're describing is the same views? It's not like it gets bounced around. It's not randomized. This time you get 30%, meaning pages one through 30 out of 100. Next time, you get 30 to 60 out of 100.

PALMA: In the Google environment, in the publisher environment, it depends upon the search. In other words, if I'm searching within in the book for a particular term, the first page I'm going to see contains that term. It may be on the half title page. It may be actually on page 136. And from there, they can browse 30% of the book.

In the situation I just talked about with Borders, it's a straight browse. They're always going to land on the first page, because there's no search.

Now we get into some realm of Google that I'm not an expert in but I know a little bit about. But I think all these other products and tools will come to bear and are starting to come to bear on book publishers and authors and how books are surfaced on the Web today and how they will be surfaced tomorrow.

If you know anything at all about search algorithms – and you probably don't – there's some basic stuff that the whole world knows, and that's the way to increase your relevancy rankings, whether it be a Web page itself or the actual site itself.

The way to do that is to ensure that other people are linking to your site or to your book, in this case. So inlinks are actually really, really important to relevancy rankings.

Google actually was started by the founders when they were grad students at Stanford, and the kind of thought experiment that was Google's little acorn was if we could possibly index all the books in the Stanford Library and surface them as one would Web pages, how would we determine relevancy? And there are about 150 to 200 different factors of relevancy.

But what they hit upon that would be really, really important, they stole from journals, academic journals. And that's the idea of citation.

So if I'm a graduate student, I write my paper, I get it published, and it's cited by 30 other people, but those people are all poor starving graduate students like myself and I'm cited in their first papers, yeah, 30 citations is nice, but those aren't very weighty citations. But what if those 30 citations came from the 30 top people in my field?

And then we could trace back from there a tree. How many people cited those guys? And the guys that cited those guys, how many people cited? And that massive calculation could be brought to bear on a single paper. You could surface relevancy in that way.

That's precisely how we surface Web pages today. It's called page rank. It's not the sole way we surface relevancy in the rankings, but it's one of the key ones.

So anywhere where we can point to a book, surface a book, point back to the publisher's website page for your book, that's going to raise the relevancy ranking of that page where somebody buys a book.

How many people have blogs? Anybody write blogs? A few people? Have decent traffic to your blogs in general? Yeah? Good.

Here's obviously just a small way that you could actually encourage links from other sites. Obviously, you want to keep your blog material as fresh as possible because one of the ways that Google ranks stuff on the Web is how relevant it is. So if it's the same content that was there two months ago, it's going to start to get buried in the relevancy rankings. So blogs should be kept fresh and be kept active.

But you have the opportunity to do all kinds of things in blogs. Here's Chris Anderson's blog, which, by the way, he probably makes more money from AdSense, Google AdSense program, running ads against this blog than he does from his book, and his book was a bestseller for many years.

But it gives you the opportunity to actually in essence do a little bit of shameless cross-promotion so that people are, if they're interested in your publications, can actually find them. Every one of those links goes back to a product page of some sort. For some bloggers, it's Amazon's page. For many other bloggers, it's starting to become Google Book Search and the book on Google Book Search. For others, it's the publisher's product page.

But the idea is that other people that come to your blog and contribute to your blog may be other bloggers, and they're actually going to point to your blog or a piece of your blog and actually link to it on their blogs. So you have this whole ecosystem, this thought ecosystem, within your particular discipline and you're creating and enmeshing your ideas and your books into that ecosystem.

And the more people that sort of pile on, the better your rankings are going to be just in regular old Google organic search, and it's also going to help point people to the books in Book Search as well.

There are lots of different tools out there. I happen to know about Blogger because it's a Google product, but you can set up a blog in about 15 minutes. You have templates. You choose a template, username and pass, and you're off to the races pulling in information, setting up links and most importantly, expressing and sharing your ideas with other people in your field. It's not a hard thing.

What I actually didn't know anything about, because this really gets into the back end, but it's pretty important, is that there are actually analytic tools that you can use to actually see what traffic is coming to your blog and what people are doing once they get there. In other words, what words are they most linking to? What are they doing after they get to the blog? Are they linking on certain links that you put there more than other links?

So you can actually use these tools that are free and part of most blog utilities have these things, and actually see exactly what people are doing when they get to your blog and what they're doing after they leave.

Something called FeedBurner – there are other things like it out there – help you to syndicate your ideas, in essence. So if I have –

I like the Freakonomics blog. Has anybody ever read the Freakonomics book or see their blog? It's now a *New York Times* blog, by the way. But I want to know, anytime Steven Dubner says something, I want to know about it, so I get an RSS feed and it populates my reader in Google every time he posts something new to his blog. Because who has time to go out on the Web and remember to tap into all the people that you care about hearing or reading?

So you to go FeedBurner and you actually kind of in essence register your blog and it makes it work interoperably with RSS feeds of all kinds. It makes sure that anybody could embed your blog anywhere, and anybody could set up a feed of your blog anywhere. And those are tools that are completely free for you to use.

And you now have the sum total of my knowledge of blogs and FeedBurner.

Does anybody use Google Alerts? I just think it's the coolest thing ever. You go into Google and you set up an alert and you set up your alert that says Chris Kenneally. Any time Chris Kenneally is mentioned on the Web, a little e-mail pops open into my box with a link to it and I can go and see what was said about.

Now, I use it extensively for particular publishers, for particular – the Australian Publishers Association. Turmoil, constantly. I want to know any time something happens on the Web related to them. I want to know about it. This will automatically populate that.

M: You're trying to put clipping services out of business?

PALMA: (laughter) Well, clipping services are offline, so to the extent that there are clippings to be clipped, that's not what we do. But if it happens on the Web, it is indeed what we do.

But from a marketing standpoint, I think it's pretty important. Obviously, you can track your own books and you can track who is mentioning your books and/or yourself on the Web. And I just find it extremely useful.

From the publishing perspective – and I'm amazed that I talk to so many publicists from my old life who go, trends? Who? What? Trends? They find out through the clipping service two months later that a book was actually starting to pop somewhere for whatever reason because something happened in the world, but they never knew about it until the clipping service gave them the clippings. They could know in real time.

M: Can I just back you up on that?

PALMA: Please.

M: It is incredible how quickly (inaudible) when somebody posts something that hour (inaudible).

PALMA: That's because you're showing up in places like the *New York Times* and CBS.com.

M: (inaudible) for anything you choose. In search results (inaudible) for something that appears (inaudible) yesterday or (inaudible).

PALMA: Yeah. The only exception to that is if it appeared in a very, very obscure place that the Google crawler doesn't index very often. So in other words, most sites get indexed and updated pretty regularly. But let's say it's my wiki from me and my brother for our golf scores. Not that I play golf anymore, but let's say that we had a wiki and we were tracking our golf game. The Google crawler's probably not going to come back to that on an hourly or daily basis. But it may come back to it on a two-week basis or a 30-day basis, so you may see some older stuff in there. But for the most part, it's literally stuff that's happening within the last 24 hours.

Any questions about that?

So YouTube. One of the fastest-growing if not the fastest-growing website on the planet right now. Nobody can actually keep up with it.

What's happening trend-wise is that the users are getting older. This began as really an 18- to 24-year-old phenomenon when it was goofy, crazy, short-form content. It's now being used in all walks of life and all industries. The number of people that downloaded videos last month numbered in the hundreds of millions worldwide. It's a major, major phenom.

Nobody's figured out how to make any money on it yet, but it's a major phenomenon and a huge marketing phenomenon as well. So because of that, publishers who are definitely, definitely videophobic are actually starting to look at this as a means of marketing, as are authors.

There are obvious trade author applications here. If you're a sort of name brand author, chances are you're going to be able to post something up here that's going to be relevant to your book and it's again, another way to raise awareness of the book, of the author and of the brand.

But I think it goes well beyond trade books as well. Today, if you give lectures and they're filmed and you can get your hands on copies, it certainly helps to put those lectures up on YouTube, and book signings, etc., that makes sense.

But what I hadn't realized is that there's this huge movement under way for whole universities to develop YouTube channels and to begin to create video content surrounding every aspect of campus life as well. Berkeley has a particularly good one. We just signed them recently.

And I noticed on the Berkeley site that you could actually gain access to pieces of lectures, if not whole lectures, from almost every course that they run. So it's an opportunity for the prof to actually get before a YouTube audience and get their ideas out there in the world as well to go well beyond the idea of text in order to broadcast their selves, basically, out into the world.

And you'll notice down here that there are – again, the social Web is absolutely key to the phenomenon of YouTube. Videos are posted on all kinds of other websites. I would say fewer than 30% of the videos that are viewed on YouTube are people coming to YouTube. They find them other places and there's this huge viral component because it's so easy to actually forward a link to a YouTube video to a hundred of your friends. And this stuff happens all day, every day, every second.

Now we get into some real futuristic things, and not necessarily far off into the future, but we haven't launched these things yet, so you're getting a sneak peek into a couple things that I think also might be pretty important.

With the exception of this. This actually exists today, but it's setting up my next slide. Today, if you have a Gmail account and you go to a book in Google, you can sign in. And when you sign in, you can actually develop a library, so you can actually put books that you come across in Google Book Search into a library and they would show up with book jackets and you could actually share that library with other people.

So say that you want your seminar students to look at five different books that you think are really important to a particular topic they're covering, you could actually share that library with them. They could go and preview those books in Book Search and then find them in the library from there. But it's an easy way to actually generate sort of a community aspect to Book Search itself.

And people can actually write reviews and tag those things and do all kinds of things that you'd expect you'd be able to do in the social network of the Web.

But what's going to get really interesting is when we widgetize this thing. The big buzz word on the Web is widgets and gadgets. We call them gadgets for some reason. Most of the rest of the world calls them widgets.

But the idea is that you would have a tool, a very easy tool, for somebody to actually put that library or that singular book anywhere on the social Web. So you think about Facebook, think about your Blogger site, think about a wiki, think about almost any place on the Web that you'd want to display a book and talk about a book, you could actually put a link – or will be able to put a link – to Book Search. They'll actually display the book jacket as it appears on our site and click somebody through to get a preview of that book within Book Search.

So what's really going to be important is that that becomes interoperable. In other words, somebody is on LinkedIn and they have a profile on LinkedIn, they want to be able to put that book up there, but they also have a Facebook account and they want to make sure that that book also appears on their Facebook account without them having to utilize a whole new tool from Facebook that LinkedIn didn't have.

This phenomenon is called Open Social, and it's Google's effort to actually open up social websites to a single platform that will be interoperable, so that when you build these social aspects of your own sites, you can have them populated wherever you may be on the Web or wherever other people may be on the Web.

Does that make sense?

I kind of buried the lead here. How many people know about Google Knol? Heard about Google Knol? Google Knol, not yet launched, is going to fulfill a really important need on the Web. Today, Wikipedia is an extremely popular site.

I don't necessarily mean to put this up against Wikipedia because it's going to continue to be a really important site. But the thing that's lacking on Wikipedia is the idea of authority.

The author actually is invisible to Wikipedia and is not actual testable as an authority on Wikipedia, and we needed to kind of find a way to address this ourselves. So the idea of Knol is that the author would be front and foremost in the actual short-form content dealing with a particular topic, a particular short-form topic.

So there would be opportunities to post actual content itself, information about the author, and of course, all kinds of links to things like your books. So that begs the question can anybody write a knol, because if anybody can write a knol, doesn't that defeat the whole purpose of authorship?

It does if it works like Wikipedia where everybody gets in, where it's a completely democratic process and anybody can write a knol and everybody can get in and everybody can be seen. Everybody's going to be able to at some point write a knol, but that doesn't necessarily mean that that knol is ever going to surface in search results, pages one through 55. And whoever looks past page two on the search results?

So there'll be factors of the knol related to authorship, and I don't know what those are but I've got to believe they would be things like publications, papers, citations, other places that you or your books are mentioned on the Web, all kinds of things that will determine relevancy so that Joe Shmoe down the street can't write a knol on your topic and actually bounce to the top. We want the authors that already have some visibility in a particular field and some credibility in a particular field to surface as an answer to a short-form query.

So when we announced that Knol was happening, the first thing that happened was the phones rang off the hook at Book Search and all our publishing partners said, see, we knew it. You're going to be a publisher.

Well, again, we're creating a platform but we're not becoming a publisher any more than creating a platform for a blog is us becoming a publisher.

We've turned them around on this idea because we said, hey, guess what? Short-form content can point to long-form content. So anywhere where your author appears, they can promote the longer form content, the book. And they've actually captured the audience in the short form of that content. So it's actually going to be a huge promotional vehicle for publishers.

What we haven't kind of worked out yet is what's the mechanism for posting those things. We'll be launching this thing in a matter of weeks, and it's kind of invite-only. And it can't work that way. We've got to develop a platform where

we can have publishers and authors begin to actually post their own knols. I have no idea how that's going to work yet, but I know that that's the ultimate direction.

Too much text on this page but there'll be plenty of information on the Knol site to help you along with actually writing them.

Any questions about that? Yeah.

F: Not specifically about knols, but I want to go back to the whole Google search. I use them and I hope that everybody in this room would appreciate hearing these arguments that Google Books is making that in order to use (inaudible) digitize (inaudible). So we know that some publishers have not favored that argument (inaudible) Google Books that copyright (inaudible). Do you have some status on the (inaudible)?

PALMA: I can't give you status only because I don't know anything. I'm not sort of privy to it. But I do want to make the distinction between – the law suits are focused on the scanning done in libraries only, obviously. Because in the partner program, we have an agreement with the publisher or the author to do certain things with the books. We have a contract or an agreement, an online agreement, to do certain things with those books.

On the library side, we still believe we're well within the bounds of fair use as we show snippets only of those texts as those books are scanned and shown on the Web. We do not show browse pages for scanned books that are in copyright from the library program.

I'd also like to hasten to add that every one of those publishers named in the law suit are very active participants in the partner program, so we've gotten to a very enlightened place for America. They've said, we disagree with you over here but we absolutely understand the value of the partner program. And they're terrific partners and they have thousands and thousands of books in the partner program as well, every single one of them.

F: An earlier edition of my book was scanned into Google Books, and about nine months ago the fifth edition came out. Is there some way the reader knows that what they're reading on Google Books is an earlier edition?

PALMA: So it was scanned as part of the – did you put them into scan?

F: They did it, apparently.

PALMA: It's the library program.

F: (inaudible)

PALMA: Or the publisher did. OK.

F: But it's got the fourth edition scans and the fifth edition is the current edition.

PALMA: Yeah. It's really up to the publisher to actually go in there and supplant the earlier edition with the latest edition. We try to actually – part of the algorithm is going to take into account pub date, so you'd expect that while the earlier edition is still going to be there in results, that the newest edition, if it's in the program, will always be first.

F: What if the new edition isn't scanned?

PALMA: It would be in there with metadata, you would expect, if we've indexed the metadata.

F: So in other words, if I go to Google Books and I look up my book, as the author I can see a full edition but another user wouldn't know that.

PALMA: They might, as I say, because we actually have agreements with all kinds of metadata providers to fill gaps where we don't have books. So if we don't have that book yet, chances are we have the metadata that we've put in the index through Bowker. So you would expect that the scanned book would appear side by side, depending on the search, right alongside the new book. It should be apparent to the user which one is the newest.

The difference is that one would be metadata only and one would be browsable until the publisher puts in the new edition.

M: I hadn't thought about this. So (inaudible) plan. I'm sorry.

PALMA: Yeah. Sorry. Go ahead.

M: So the publisher would have to not only give you a new edition but also affirmatively say, by the way, get rid of the old one, otherwise the default is that both of them are up there?

PALMA: Yeah. We'd never want to make that call on behalf of the publisher.

M: What's the default?

PALMA: The default is the publisher has to determine which books they want to appear in the program and which books they don't want to appear in the program. So there's no way that we'd be able to scale something where we would be able to know which book to remove. Plenty of librarians would tell you they'd want to see all editions up there.

F: (inaudible) you could have material on there, available to students so they can (inaudible).

PALMA: And that's the material –

F: (inaudible) ourselves but for allowing them access to (inaudible).

PALMA: Sure. So that's literally the lecture itself.

F: It's literally (inaudible).

PALMA: Yeah, so they can – because they slept through it earlier in the semester, they can go back and actually – man, I wish I had that when I was in college.

No, as I said, it's crass commercialism, but I think it's part of the promotional process to be on the Web in any different form that you could be on the Web. It's up to you how far you want to push the envelope on that.

Any other questions? No?

Well, thanks very much. It was great to come back and talk with you all.

(applause)

END OF PRESENTATION