Nathan Rinne

Concordia University

Mar. 14, 2012

Library Technology Conference

Macalester College, St. Paul, MN.

Teaching with Google Books: research, copyright, and data mining

Short Description:

Do you know about Google Books? Join an exciting tour that will not only introduce the Google Books Project and its history, but will share ideas about using it as a springboard to delve into issues like: a) data-mining; b) copyright law; and c) research, both personal and scholarly.

Long Description:

Google's Google Books site is a rich resource that is probably underutilized by most educators. It has all kinds of potential for a) getting students into the research process in a way that they will enjoy (for example, they can see how a famous quote has been used/quoted , find out which books cite the journal article they are interested in, or check to see if a specific book covers a topic that they want to explore, etc.) ; b) teaching them about the deeper civic purpose and the evolving state of copyright law ; and, c) exploring, with the help of Google Book's Ngram viewer, the promise and ethical issues surrounding the issue of data-mining and "non-consumptive" research, or research that is accomplished by "mining" books for data, as opposed to reading them (here issues of online privacy can also be touched on, as the same kind of data-mining that is used in the Ngram viewer can also be used to produce advertising portfolios on those who read).

In addition to this, at the beginning of the presentation I will talk a bit about what Google Book search is, where it came from, and how it has evolved a bit over the years.

Intro

The title for this presentation is "Teaching with Google Books: research, copyright, and data mining". One thing the title and abstract does not tell you is that themes of "education", "freedom" and "ethics" are going to be interwoven throughout this presentation.

I'll begin by focusing on these.

In his autobiography, Benjamin Franklin wrote of the how the growth of lending libraries had both educated and democratized America:

"These Libraries," he wrote, "have improv'd the general Conversation of the Americans, made the common Tradesmen & Farmers as intelligent as most Gentlemen from other countries, and perhaps have contributed to some degree to the Stand so generally made throughout the Colonies in Defence of their Privileges."[1]

Years later, Thomas Jefferson said that "Knowledge is the common property of mankind"[2], but we know, as did Franklin, that historically much knowledge had been the privilege of the few.

By the way, when we say knowledge, what do we mean? We can talk about knowing persons, knowing about facts, and philosophical definitions like "justified true belief". But let's not get too complicated. How about this?: "knowing how things regularly transpire in the cosmos – and how these things can be understood (and perhaps harnessed) to help us move ever more successfully within it… in other words, freedom…."

This is what knowledge does. In some sense, knowledge is power because it allows for freedom. And to talk about the "liberal arts" really means to talk about the arts "suitable for a free man." They are "the areas of learning that *cultivate general intellectual ability* rather than technical or professional skills. The term liberal arts is often used as a synonym for *humanities*, although the liberal arts also include the sciences. The word liberal comes from the Latin liberalis, meaning suitable for a free man, as opposed to a slave." (italics mine)[3]

So who can open the floodgates of knowledge and education to all people that they may thrive and flourish? Who can be the liberator and champion of the people?

Is it not Google? (we'll get to Google Books specifically soon) How is it not Google?

For let *me* define "Google". Google is not only "organizing the world's information and making it universally accessible and useful". It is, practically speaking, the instant gratification of our information needs and wants. It helps us to do what we want… what we think is right… to freely pursue the goals we think we should pursue.

Does this not sound like it is related to freedom? Don't you remember the wonder you felt – the freedom – when you first used Google? Lawrence Lessing describes libraries by saying: "[They]

gave us the freedom to research, regardless of our wealth; the freedom to read, widely and technically, beyond our means. It was a way to assure that all of our culture was available and reachable…"[4]

And how is Google – especially Google books – any different?

Does Google not mean freedom? How can you doubt that it does not? After all, let me tell you about the Google Books project….

(Having fun yet? Surely I jest a bit, but hopefully in the service of making serious points…eventually)

Part I – Brief Google Book history and tour

"What is a free society obliged to do with its written words?"[5] Well, if "information wants to be free", Google has put real legs on that idea as regards books. In 2004 Google signed agreements with 5 major research libraries (Harvard, Stanford, Oxford, University of Michigan, and the New York Public Library) to scan the books in their collection for usage in Google Books. Depending on a book's copyright status, the full text would be made available freely online. At this time, Google frequently talked about creating a "universal [digital] library" – something its founders had wanted to do right from the company's beginning.[6] Since those initial contracts were signed, now 7 long years ago, they've got over 50 other such libraries to help them en route to their 2020 goal of digitizing 130 million books - the amount they estimate exist[7]. As Sergey Brin, co-founder of Google says, they "feel this is part of our core mission" – after all, "often when [Sergey does] a search, what is in a book is miles ahead of what [he] find[s] on a Web site."[8] What is the world of information without books?

http://en.wikipedia.org/wiki/File:Sergey_Brin_cropped.jpg

Surely, in this age where "everything that exists is online", "widespread digital access is key to scholar's work", right?[9] And who could question Google's "commitment to putting those digital riches within the reach of multitudes"?[10]. At the time, the government was not about to do a "Human Knowledge Project" (like the "Human Genome Project").[11] And Michigan's library head Paul Courant pointed out that there was no way that libraries alone could have done this - Google was digitizing in one week what was previously taking one year. That same university's President, Mary Sue Coleman, explained that the project was a "legal, ethical and noble endeavor that will transform our society."[12] After all "massive collections of material curated by research librarians over the centuries" was being made available.[13]

Let me pause at this time to discuss with you the wonders that Google hath wrought. Let's take a brief tour. If you do a search in Google Books – I'll do *Catholic Mass*[14] - you will pull up a results list and then notice several things. As you explore (you may have to go past the first page of results), you will see that some books have a "read" link, while others have a "preview" link. Others are simply labeled "snippet" or "no preview". The books that have the "read" link are available in full text as these works, usually published before 1923, are in the public domain. You can see a list of only these books by clicking "Free Google eBooks" in the left hand column (note that you can sort the list by date as well). You can also search in the text within each book as well, and if you do, you'll notice that it is very easy to use with lots of interesting tools (note the marked scrolling feature on the right). You can also do this with some "Preview" books, as Google has made deals with publishers so that you can see a certain amount of pages. "Snippet" views – where you can see your search terms in their immediate surroundings - are there either because a deal was made with the rights-holder for this limited amount or because no one really knows who the rights-holder is (more on this shortly). With the "no preview" book, no deal with the rights-holder has been made.
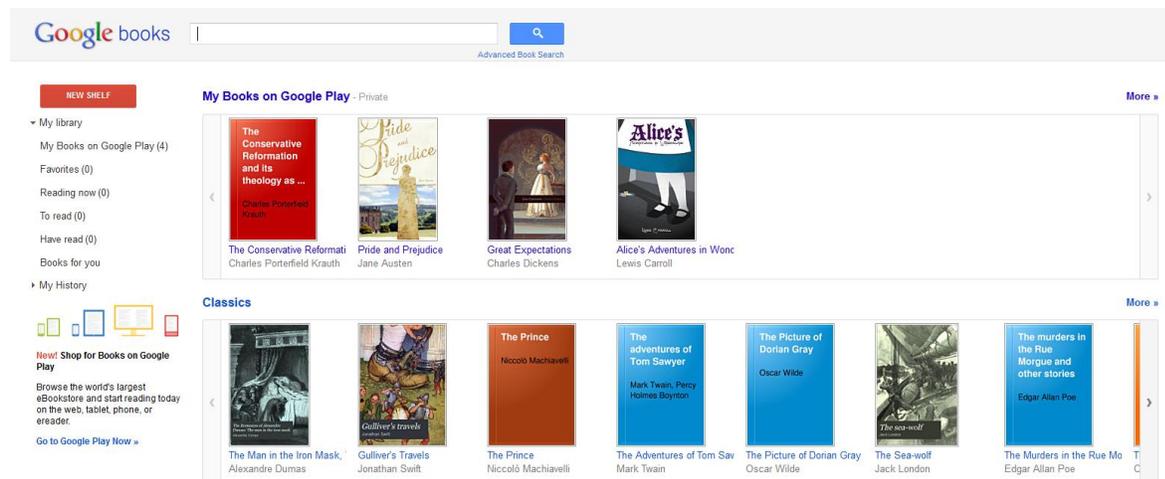
Screen shot from March, 12, 2012

Can you not feel the chains around you breaking - the light penetrating the darkness? Seriously, I must admit all of this is pretty awesome, and in full disclosure I use it all the time.

Suffice it to say, the main controversy that surrounds Google Books has to do with their scanning books that were in-copyright but out-of-print – and then providing these "snippet views". In the "eye of the storm" were the millions of particular books in this category who are "orphans" (the "vast majority" of the scanned books according to Google[15]) – meaning that no one knows who really holds the rights to these books (as the content owners have not been identified or copyright ownership is debated[16]). At the beginning of this project, Google justified their actions by appealing to the "fair use" provision of copyright law, despite lawsuits from authors and publishers. Later on though, as Google attempted to settle with its opponents, it announced that they were creating a new book business (hence the "Google eBooks" link you saw on the left-hand column)[17], where the in-copyright but out-of-print books that it scanned could be sold….

According to this settlement, Google, with the help of 125 million of their dollars, would have established an online registry to let authors and publishers register their works and get paid (63% of profits would go to them) when their titles were viewed online. The settlement also allowed persons to prevent Google from scanning their stuff and to collect money for each work that Google had already scanned by "opting out". For its part, Google would have been able to, among other things:

- show longer previews of most all of the out-of-print books (in other words, use the millions of "orphaned" works)
- allow persons to buy the books (print-on-demand or as e-books)
- allowed for ads to be shown on the book pages online
- charge subscription fees to libraries and universities in order to access the full-text of the out-of-print copyright protected works[18]

This brings us to Part II, where we will discuss the settlement some more and how this relates to education (of society), freedom, and copyright law…

Part II – Understanding Copyright Law through Google Books

Let's look at some of the pros and cons of the revised settlement[19] Google had made with the Author's Guild and Publishers, with the hope that it would be accepted. I have bolded items that especially have to do with copyright law.

Positives and negatives of the [revised] settlement[20]

Positives cited:

- Google did what librarians could have never done and what gov't likely would have never done
- **Would provide millions of people with access to millions of out-of-print but in-copyright books** (in homes, public libraries, and academic institutions)
- Could give new life to old books (not just sitting in remote storage)

- **Service would be provided free of charge on at least one terminal in all public [and academic] libraries**
- Would be adapted to the needs of the visually impaired
- Data would be available for "large scale, quantitative research of the 'non-consumptive' kind"[21]
- It would help cut down on expensive interlibrary loans – and help eliminate loans that disappoint[22]
- **"Authors and publishers [would] be able to cash in on long-neglected works"** [23]

Negatives cited:

- **Opt-out clause for rights holders of out-of-print but copyright-protected books** (why not an "opt-in" clause?)
- **Foreign authors and publishers** (U.K., Can. and Aus) **also not happy** (violations of international copyright law)[24]
- **Even though others could scan library collections like Google did, Google would have exclusive protection vs. legal action by any rights holders who might come forth** (**who is the owner here?** Is this not an effective monopoly?)
- **Is the author's guild (8,000 people) truly representative of all authors** (6,800 authors opted out)?[25]
- **Many academics want their books to be free on Google Book Search, so their ideas can be spread** (no "Creative Commons" option)
- User privacy concerns (more on this later)[26]

Other interesting questions and opinions:

- "Since when is expediency one of the core values of librarianship?"—Siva Vaidhyanathan[27]
- "Books and other printed material would quickly reach obsolescence if not easily accessible through digital technology"[28]
- Google says the "opt in" structure would not be viable[29]
- "**Clarifying the copyright status of millions of digital books, the deal would also make it easier for firms other than Google to strike deals to use them**" (which Google wanted to allow)[30] [31]

All kinds of players –librarians, scholars, publishers, authors, techies – got involved in the commentary and debate. Before the ruling on the settlement, the *Economist* said: "The case has stirred up passions, conflict and conspiracy theories worthy of a literary blockbuster."[32] So what happened?

In March of 2011, a federal judge threw out the 2008 settlement that Google had reached with author and publisher groups to make millions of out-of-print books available online. Judge Chin

said: "[A] Universal digital library would benefit many… [but the] current pact would 'simply go too far'"[33] The six negatives previously cited all found their way into the 48 pg. decision…. In short, the settlement would have in effect rewritten U.S. Copyright Law, which the judge said should be decided by Congress.

Currently, the settlement between Google, the Authors Guild, and the Association of American Publishers can now be revised (though it would almost certainly need to be "opt in"), but while Google reports of some progress made with the publishers[34], the Author's Guild is not only renewing their original lawsuit (that prompted the settlement), but is also going after HathiTrust (pronounced "hot-ee"), the consortium of university libraries that, due to their agreements with Google, received individual copies of the books they scanned. Their complaint states: "by digitizing, archiving, copying and now publishing the copyrighted works without the authorization of those works' rights holders, the universities are engaging in one of the largest copyright infringements in history."[35] Yet Hathi Trust, for their part, is not even showing "snippet" views, but only telling persons on which pages their search terms appear.

Google, in turn, is now saying that the claims of the Author's Guild and others should be dismissed, since they themselves do not own copyrights (only representing some of those who do).[36] They have also experienced another setback as the Supreme Court recently ruled that some items that had been in the public domain – items that Google had scanned – can now be taken out of it. In other words, "copyright restoration."[37]

Nevertheless, they are still scanning.

Google Books: the perils

The fact that the settlement was rejected does not mean that it did not have many positive features, as many respected commentators pointed out. However, the idea of letting a powerful company change the copyright landscape through the settlement without any government oversight or legislation was simply too much.[38] A man named Santosh Desai wrote about the "dangers of placing all our information eggs in a private basket".[39] In other words, there were concerns that this would not lead to freedom, but bondage. Before the court ruled against the settlement, Harvard library director Robert Darnton, noting how price-gouging academic journals once had been produced "solely in the spirit of free inquiry", said, "Google's record suggests that it will not abuse its double-barreled fiscal-legal power… But what will happen if its current leaders sell the company or retire?"[40] Copyright crusader and Harvard professor Lawrence Lessing echoes that concern when he says, "we cannot rely upon special favors granted by private companies (and quasi-monopoly collecting societies) to define our access to culture, even if the favors are generous, at least at the start."[41]

Obviously, with its reputation, Google has an incentive to "not be evil" as they say, but as Darnton points out, "Google's primary responsibility is to make money for its shareholders. Libraries exist to get books to readers…"[42] Lessing, in his essay *For the Love of Culture*, talked

persuasively about how "*It is the environment for culture that the settlement will cement. For it practically guarantees that we will repeat the cultural-environmental errors of our past, by now turning books into documentary film [where each clip must be purchased and re-purchased]….the deal constructs a world in which control can be exercised at the level of a page, and maybe even a quote. It is a world in which every bit, every published word, could be licensed*" (italics mine).[43] For his part, Darnton notes how Lewis Hyde, in his recent book *Common as Air*, says "an enclosure movement is threatening to destroy our cultural commons, the world of knowledge that belongs to us all."[44] [45] Asking whether or not we want to "commercialize access to knowledge", Darnton sounds like Lessing when he speaks of the need for a "new ecology – one based on the public good instead of private gain."[46]

In addition, if the settlement had been approved, this would have done nothing to clarify the issues surrounding copyright law and the fair use provision of it.[47] Most people who looked at things from this perspective thought that Google, had they not pursued the settlement, would have been victorious in court, as their project was "sufficiently transformative" to be fair use.[48] Darnton thought they should have made a robust case for 'fair use' and tried to set a legal precedent."[49]

At his point, let's ask some tough questions about copyright law and the idea of intellectual property. Copyright law was originally intended to provide a balance between ownership and the kinds of use that would benefit progress in the arts and science, and it was revolutionary in its day. Obviously, as regards ownership, "Creativity requires stability", says Kevin Mattson, "You can't express yourself— write the book or article or teach the class—if you constantly worry about the next source of income… It's about having time to reflect and think".[50] In our way of thinking, this kind of ownership goes hand in hand with people having the right to be paid for their work, and if they are not, trust in society decays. The Indian Christian writer Vishal

Mangalwadi talks about how shocked he was to find Americans in Europe who had were trying to convince him that he was foolish to pay for his bus rides, since they had been in the country for two weeks and had never been asked to show their tickets. Vishal lecutred them for this, telling them how their actions were partly responsible for the cultural decay that would eventually result in the same kind of systematic corruption that plagued his native country (where bribes are necessary to pay of all the people who are hired precisely to do things like make sure people pay for their tickets).[51]

On the other hand, we realize that it is good that there are things that we share in common, so that all may benefit. As regards knowledge, we want much of this to be publically available, so that we might encourage intellectual growth, discussion, creativity, and the like. As Lessing points out, the free access that this [pre-commodification] world created is an essential part of how we passed our culture along.[52] There is no doubt that when we do think along these lines, matters can become somewhat complicated. In fact, sometimes, we come across situations where it simply does not seem right that we have such strict intellectual property laws – such as a case, for example, where drug companies push for stricter patent protections in poor countries where many people suffer from the very diseases the drugs are meant to counter[53] – but we also know that it is also not right to insist that persons who do good work should not be rightly rewarded for their labor. I think as librarians we must ask ourselves some difficult questions here. Is it right, for example, to think that in this digital world our abilities to loan to people should not be limited by place at all? Some of the libraries affiliated with the Open Library are allowing loans to borrowers from all around the world. Is place no longer a factor at all? Are they really only doing "what libraries have always done"?[54] Is it possible that things like this are in part responsible for the recent stalemates between publishers and libraries as regards e-Books?[55]

In sum, there are no easy answers, but I think that E.F. Schumacher can at least give us an idea about how to think about these things:

> "Justice is a denial of mercy, and mercy is a denial of justice. Only a higher force can reconcile these opposites: wisdom. The problem cannot be solved, but wisdom can transcend it. Similarly, societies need stability *and* change, tradition *and* innovation, public interest *and* private interest, planning *and* laissez-faire, order *and* freedom, growth *and* decay. Everywhere society's health depends on the simultaneous pursuit of mutually opposed activities or aims. The adoption of a final solution means a kind of death sentence for man's humanity and spells either cruelty or dissolution, generally both… Divergent problems offend the logical mind."[56]

Part of this balance is not only simultaneously affirming that persons deserve to be paid for their labor and "free use" should exist, but also acknowledging that some situations are simply economically unfair. In part due to his knowledge of the price-gouging happening in academic

publishing (where scholars and researchers'work, often paid for by public monies, is given freely to journal publishers who then proceed to sell their work in journals set at exorbitant prices), Darnton also saw the situation with Google as one worthy of real concern, which is why he is now pushing for the Digital Public Library of America, which I will now mention briefly.

DPLA

Instead of charging subscription fees to libraries and universities to access the full-text of the out-of-print copyright protected works, the Digital Public Library of America (DPLA) is currently the top alternative to this. Darnton has been the organizing force of a movement of librarians, academics and computer programmers who want to make not only digitized out-of-copyright works available, but many of the out-of-print but in-copyright books as well (fairly compensating the authors). Right now, its efforts are supported by grants and foundations but ultimately, a sustainable action plan will need to be realized, and federal legislation will need to be involved in this process (as regards compensating authors when people read their works online, I imagine this would occur through government monies, though I don't think such a venture in today's climate will be too successful…).Google itself has said that they are receptive to this – they have already contributed several book scans to a similar pan-European project, *Europa.* [57] As Darnton has pointed out, if Google does indeed contribute to this project and gets appropriate recognition for it, this can only be good for their reputation. [58]

To close out this section, not much changed for users of Google Books with the rejection of the settlement. Most out-of-print books are available as a "snippet view".[59] Of course, another two million [mostly in-print] books can be viewed in longer previews, due to agreements between Google and "tens of thousands of publishers that were separate from the legal settlement".[60] At this point, let's talk a bit more about the promises – and yes perils – that this project brings - for research

Part III - Google Books and Research: the perks and pitfalls

Thomas Jefferson, President of the United States and noted bibliophile, sold 6,700 books to the Library of Congress, only to get an additional 1,600 volumes during the last nine years of his life. In 1829, in order to pay his debts, his "retirement library" was sold at an auction. The Thomas Jefferson Libraries project at Monticello in Virginia has been trying to reconstruct this library and recently, a search on Google books revealed an 1880 article in the *Harvard Register*, which recorded the gift donation (of books) of a man who had purchased many of Jefferson's books to a founder of the Washington University in St. Louis. Washington University at this point discovered it owned 74 volumes that belonged to Jefferson and gave them to the retirement library, which now has been virtually reconstructed.[61]

How does Google do it?

We know that for their regular search engine, Google ranks pages largely by their popularity on the internet. In Google Book Search, their search algorithm now "….takes into account more than 100 'signals,' individual data categories that Google statistically integrates to rank your results. When you search for a book, Google Books doesn't just look at word frequency or how closely your query matches the title of a book. They now take into account web search frequency, recent book sales, the number of libraries that hold the title, and how often an older book has been reprinted."[62] We can also say with some confidence that a search in Google Book search is not only searching the books data "generated by optical character recognition (OCR) technology, but metadata incorporated from a variety of sources and possibly machine-generated as well."[63]

And what can Google do?

Google Books can do quite a lot. One newspaper columnist writes of her love:

> "I'm going to reveal one of my secrets.
>
> When it comes to historical research, my absolute favorite tool is books.google.com.
>
> Need Dutch oven history (my column two weeks ago)? It's there.
>
> Need first-person accounts of the Second Seminole War from books published in the 1850s? They're there, too.
>
> And that book that's on my shelf that I really needed because I'm on deadline and don't have time to run home and get it? Occasionally, I get lucky with those, as well.
>
> From slave narratives to old travel guides to specialized encyclopedias, Google Books can be a fantastic tool for the historian or genealogist who is short on time to run to the library."[64]

As does an academic:

> **"**Google has offered up another new toy for analyzing texts. I've been growing in my appreciation for Google Books and its controversial program of scanning in old books because it makes available, sometimes in pdf form, out-of-print texts which are not easy to locate. So now when doing research it is quite easy to track down footnotes, whereas in the past one had to copy the reference down, trudge over to the library, fill out an ILL slip, hope our librarians found a library willing to lend a 150 year old book, and then wait for it to arrive. Instead of weeks of hoping to get a glimpse of a page, now often you can find things instantly, delivered right to your desktop. (No, I don't get paid by Google for my posts)."[65]

There can be no doubt, that as Millie Jackson says, "the transformation of the way we work as scholars and researchers is tremendous". How can we not cry out "Freedom!"?

Here's a list of just some of the things that you can do….

- check to see if a specific book covers something you're interested in
- find out which books cite the journal article you are interested in
- cut down on interlibrary loan usage
- discover rare texts and those with small print runs
- highly granular searching: easily find historical concepts that are not easily located using simple library subject headings.
- confirm a quotation[66] or see how a famous quote has been used
- discover unknown authors' and works….
- and of course… access to stuff that previously only libraries had… (picking out the "best of the best" – decades of collection development work by top-ranked libraries…)

Let's look at a few of these in a bit more detail….

Check to see if a specific book covers something you're interested in: You can use "search within the book" to find words, phrases or subjects in the book to see if the book will be useful… will this book assist me in my research or collection building? Are you interested in your family history? Break out those old family trees and enter some names into that famous search box. Google Books contains genealogy books and historical books and many are "out of copyright" - which means they can be freely downloaded to your computer. Plus, a successful "vanity search" in Google Books is more rewarding than your regular web search: who doesn't want to be in an actual book?

Find out which books cite the journal article you are interested in: Are you curious to know whether or not a particular journal article – old or new – was cited and commented on in any books? It is very easy to do this with Google Books – just quickly throw in the author's last name and the title of the article and periodical. And one of the nice things about Google's algorithmic search is that it allows for far more mistakes in your original keyword search, so you can even be a little sloppy here.

Cut down on interlibrary loan usage: Sometimes faculty have some rather "esoteric research" and here Google Books can "assist with ILL requests which sometimes can not be filled because of the age or rarity of a work." Although it may take some fine-tuned advanced searching, it is possible that you might be able to get a PDF text in the patron's hands in very little time – easily and inexpensively fulfilling a request that otherwise would have not been possible.[67]

Discover rare texts and those with small print runs: One newspaper columnist writes

"If you want information on the history of an area where an ancestor lived, type something like "History of Pike County, Illinois." When I entered that term, I was shocked to learn that an 1880 book with the same name has been digitized and is available for free download at Google Books. I had read the fragile book about four years ago when my husband and I stopped at the library in Pike County. Reading an old book that has digitized by OCG is even better than reading it at a library because one can easily search the book for people, localities and other key words." Later, she writes, "Perhaps someone in your family was a minister or helped found an early church. When I entered the term 'Baptists in Missouri,' I learned that an 1882 book, 'A History of the Baptists in Missouri,' has been digitized and is in public domain."[68]

Highly granular searching: easily find historical concepts that are not easily located using simple library subject headings: For example, Millie Jackson shares how she was able to search for the concept of "pin money" which in the 18th century was a term used to describe spending money that women had. When she did the search in WorldCat (back in 2008), she could not find the term in any book records. "Pin money" was not a subject heading, nor did it have a "see also heading", nor did it appear in any book's table of contents that had been entered into records. On the other hand, Google Book search quickly located several thousand results from the earliest appearances of the terms upward on.[69]

The freedom is palpable, is it not?

But what can't Google do?

- No authority control
- OCR without human help (more on this next part)
- Flawed dates
- classification errors
- mismatched titles and authors
- Gov doc issues, multi-volume issues, scanning errors

There are some things that bind our research as well with Google Books. Let's look at these in more detail:

No authority control: In the past, if you wanted to have all the works of a certain author, you would need to look through a wide variety of indexes and abstracts in order to get all of their works. With Google Book search you may also need to search under a number of variants of an author's name in order to locate all of their works… and you may also need to search through thousands more results.[70] In library catalogs, this can be done quickly as regards books, for example searches for "Currer Bell" will retrieve books by "Charlotte Bronte". In addition to

this, Google Book Search does not appear to utilize features that most library catalogs would offer, such as cross references and see-also references.[71]

OCR issues: The linguist Geoffrey Nunberg writes:

> "Of course, there are bound to be occasional howlers in a corpus as extensive as Google's book search, but these errors are endemic. A search on "Internet" in books published before 1950 produces 527 results; "Medicare" for the same period gets almost 1,600. Or you can simply enter the names of famous writers or public figures and restrict your search to works published before the year of their birth. "Charles Dickens" turns up 182 results for publications before 1812, the vast majority of them referring to the writer. The same type of search turns up 81 hits for Rudyard Kipling, 115 for Greta Garbo, 325 for Woody Allen, and 29 for Barack Obama." (Or maybe that was another Barack Obama.)"[72]

Even in a search Google itself drew attention to for their n-gram viewer (reviewed in the next section) almost a year and a half ago, one can see how there are still hundreds of references to "Abraham Lincoln" in the early 1800s.[73]

Flawed dates: As Marc Parry writes "Worrisome questions remain about the quality of Google's data, which may be less like the library of Alexandria and more like a haphazardly organized used-book shop".[74] Again, Geoffrey Nunberg pointed out a number of howlers. For example, you may be surprised to hear that hundreds of modern books were actually published in 1899, including Stephen King's *Christine* and Robert Shelton's biography of Bob Dylan. These dates have since been fixed, but if you spend some time in Google Books, it won't take you long to find more.[75]

Classification errors: Google claims that they are not responsible for most of the errors having to do with metadata (rather it is their library and publishing partners who are to blame) but as Nunberg argued in his article, this explanation does not conform with the evidence, and this is nowhere more clear as with classification. After initially declining their use, Google eventually decided to use truncated Library of Congress subject headings…[76] and mixed them with the BISAC headings used by publishers. If BISAC categories had not been applied to a book by the providers, Google assigned one, and Jon Orwant, who manages the Google Books metadata team, estimates that they got this right about 90% of the time. The other 10% leaves us with H.L. Mencken's *The American Language* as Family & Relationships, an edition of *Moby Dick* labeled Computers, *The Cat Lover's Book of Fascinating Facts* as Technology & Engineering, and Susan Bordo's *Unbearable Weight: Feminism, Western Culture, and the Body* (misdated 1899) as Health & Fitness. Nunberg has lots of fun talking about Google's "fine algorithmic hand …." and getting in jabs like the following: "Google has taken a group of the world's great research collections and returned them in the form of a suburban-mall bookstore".[77]

Mismatched titles and authors etc.: These can make for some real entertainment as well. *Madame Bovary* by Henry James. *The Mosaic Navigator: the essential guide to the Internet interface* by Sigmund Freud and Katherine Jones.[78] To close in understatement by Marc Parry: "Google Books… isn't marked with rigorous metadata, a term for information about each book, like author, date, and genre."[79]

Bibliographic information

Title	Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Ever
Author	David Weinberger
Publisher	Basic Books, 2012
ISBN	0465021425, 9780465021420
Length	256 pages
Subjects	Computers › Social Aspects › General

Computers / Information Technology
Computers / Social Aspects / General
Information technology
Information technology - Social aspects
Information technology/ Social aspects
Internet
Internet - Social aspects
Internet/ Social aspects
Knowledge, Sociology of
Social Science / Sociology / General
Technology & Engineering / Engineering (General)
Technology & Engineering / General

Bibliographic information in Google Books utilizing LCSH and BISAC

Gov doc issues, multi-volume issues, scanning errors: While U.S. government documents should all be freely available since they were never under copyright, many of these are available only in snippet view or with no preview at all.   Duguid noted Google Books' difficulty "in identifying volume number in multi-volume works".[80] And finally, many have noted that all manner of interesting scanning errors – from hands to blurred pages – can be seen in several of the books Google has scanned.[81]

Other concerns:

Another concern about Google Book Search as regards research might be that "quick and dirty" becomes the "one ring to rule them all" (for the general public as well as scholars), although the content on Google Books would no doubt be preferable to that of the web in general.   As Chris Thompson says, summing up Geoffrey Nunberg's view, "because Google has an effective monopoly on the world's only digital archive, researchers will come to depend on it, erroneously assuming that Google's got the details right." Nunberg says: "Of course people will use it instead of their local library. Who wouldn't? I use it all the time".[82] After all, "if it's not on the web it doesn't exist" – but fortunately now it does! How likely is it that most people will be able to exercise constant vigilance when it comes to being aware of their information options? The more rights to full text Google is able to secure, the less necessary (and perhaps more invisible) that "find it in a library" link may become…..

http://upload.wikimedia.org/wikipedia/commons/b/b7/Unico_Anello.png

Alternatives?

In the end, while it is important to remember how Google is "constantly tweaking the company's search engine in an elusive quest to close the gap between often and always", as Saul Hansell puts it, there is also the need to discern when "to Google or not to Google".[83] As regards many scholars, of course Google Book search will always be insufficient by itself. This is not only because of the quality of scans and lack of in-text metadata that can be used to search the full text (which will also be touched on in the next section), but also because of the fact that Google's uptake of the library data that many scholars have come to rely upon has been haphazard and inconsistent. Scholars want to be able to quickly locate multi-volume sets, be able to quickly distinguish between various editions (and the year they were printed), and be able to count on accurate classification, subject headings and authority control functions – even if they might currently not realize how useful this underlying "structured data" is in helping them get useful materials (especially when doing research in a field that is not their specialty).

In this case, one can't do better (for now) than to supplement Google Books with Hathi Trust, the consortium of over 60 academic libraries that are using the Google scans and others to form a massive online library catalog complete with full text. Hathi may not have the "secret sauce" of Google's search algorithms (that sometimes are very useful, and allow for some misspelling and correction functions), but they are doing their best not only to maintain the things mentioned above (accurate classification, good subject headings and more authority control), but to also take the additional step of determining which orphan works can be in the public domain by trying to locate the copyright holders individually. Last time I checked, less than a year ago, Hathi had tried to contact some over one hundred thousand rights-holders, and found out that more than ½ of the orphans could be put in the public domain. Again, last time I checked, there were several post-1923 books that were in Hathi but not Google Books. A non-profit, Haithi's

main charge is to "introduce trustworthy curation and permanence for the cultural record".[84] It is for reasons such as these that some hope that the Hathi Trust, along with other depositories like Project Guttenberg and the Internet Archive, may soon be able to play a part in the dream of the Digital Public Library of America addressed in the last section.

Finally, Hathi Trust, also has a real concern for the quality of digitization – and one of the reasons for this concern has to do with the fact that Hathi Trust wants to "serve as a body of content for large-scale computational needs". So does Google of course, and this brings us to our last section.

Part IV - Google Books and the Digital Humanities

Have you heard of the "Republic of Letters", the project which attempts to map the correspondence of the primary movers and shakers of the Enlightenment?

Here's a brief clip:[85]

Picture from: http://artsbeat.blogs.nytimes.com/2010/11/16/digitally-mapping-the-republic-of-letters/

Using algorithms to "free" the data!

It seems that the freedom we have been mentioning off and on in this presentation is reaching a fever pitch with this last topic of discussion…

Computer programmers can certainly do some impressive stuff with computers. Granted, a project like this involved a lot more than scanning, OCR, and fancy algorithms to extract relevant data. Lots of human attention – including a ton of structured data, or metadata[86], was involved.   Nevertheless, Google, with their Books project and algorithmic expertise, has been able to get in on this "digital humanities" action in a big way.   As Jon Orwant, engineering

manager for Google Books put it after attending conferences about data mining in the humanities, "I realized…we were sitting on this huge trove of value".[87]And since, according to Siva Vaidhyanathan digital humanities research is "full of great ideas and short on the tools needed to execute these great ideas, largely because of a lack of money"[88], once again, Google comes to the rescue!

How did this play out? Eric Lieberman Aiden, a recent graduate of Harvard University, and Jean-Baptiste Michel, a postdoctoral fellow at Harvard, worked with Google to improve the their Book search dataset (5.2 million of its books from 1800 to 2000, or about 4% of all books ever published[89]) and have created a project that shows how vast digital databases could be of use not only to scholars, but others as well. As Lieberman Aiden says "The goal is to give an 8-year old the ability to browse cultural trends throughout history, as recorded in books".[90] In short though, they are applying "high-turbo analysis to questions in the humanities", and they call these efforts "culturonomics"[91] – evidently after "genomics" (Lieberman Aiden's expertise is in applied mathematics and genomics).[92] Google unveiled the software tool they helped build on the 16th of December in 2010 to coincide with the publication of a paper introducing culturonomics by Lieberman Aiden, Michel, and 10 other non-humanities colleagues in the journal *Science*.[93]

Jean-Baptiste Michel and Eric Lieberman Aiden presenting.

The tool they created is called the "Ngram viewer", and it examines "the occurrence of n-grams" which are "sequences of a certain number of words, this number denoted by n."[94]   In short, it lets you see, over time, the frequency of words in books, and the "periods of time are statistically evened out, so that the far greater number of titles available in the last few decades does not overwhelm the relatively fewer number of titles in earlier centuries."[95] To get a taste of what the tool does, we can take a very brief tour here.[96] We can also try our own searches – for example, check out how the decline of "propaganda" goes hand in hand with the rise of "Orwellian", or how "depression" overtakes "melancholy".[97] It doesn't seem a stretch to me to say that teachers who want to teach critical thinking not only about these concepts in history but the changing

nature of language itself have a gold mine here (hopefully while also talking about the issues with OCR scanning as well: as Pope and Holley pointed out, their Google Book search for "porno indians" netted many results, that is the "pomo" tribe[98]).

But these are not the only things these new trailblazers would like to show us. As the official Google blog touts: "[Leiberman Aiden's and Michel's] work provides several examples of how quantitative methods can provide insights into topics as diverse as the spread of innovations, the effects of youth and profession on fame, and trends in censorship."[99] And more folks are in on the action. A couple of the most prominent are Franco Moretti, a Stanford professor of English and comparative literature, and Matthew Jockers, co-director of the Stanford "Literature Lab". [100]

Moretti, noting how much of literary history consists of "anecdotal studies"[101] (for example, " 20,000 or 30,000 novels released in the "19-century British heyday of Dickens and Austen" are never studied[102]), talks about how important it is to use "methods from linguistics and statistical analysis". So, his "Stanford team takes the Hardys and the Austens, the Thackerays and the Trollopes, and tosses their masterpieces into a database that contains hundreds of lesser novels. Then they cast giant digital nets into that megapot of words, trawling around like intelligence agents hunting for patterns in the chatter of terrorists.

"Learning the algorithms that stitch together those nets is not typically part of an undergraduate English education…."[103]

And there are many other scholars joining the ranks as well, using the Google Books dataset to "mine the text", otherwise known as performing "non-consumptive" research, or "distance reading" (which means no real reading at all). In the late summer of 2010 Google, eager to "highlight the scholarly value of Google's achievement"[104], gave its "Digital Humanities Awards" along with $1 million dollars over two years to a dozen winners – teams of English profs, historians, biographers and others.[105] (Google has identified eight "disciplines of interest" for its program: literature, linguistics, history, classics, philosophy, sociology, archaeology, and anthropology[106])

Here is a sampling of some of the findings of research of this type. One can:

- Trace the novel going from an aristocratic literary form to a more popular one: First names like "Jim" do not appear before the 1870s, whereas before there were many "Mr. Knightleys" and such.[107]
- "Calculate the rates at which irregular English verbs became regular over the past two centuries"[108] – "chid" and "chode" went to "chided" in only 200 years (the "fastest verb to regularize")[109]
- "Detect the suppression of the names of artists and intellectual books published in Nazi Germany, the Stalinist Soviet Union, and contemporary China" [110]

- Realize that writing in a specific literary genre is "immediately restrictive of artistic freedom in ways writers never would guess" – The "place-centered" genre of Gothic novels "(think: castles, dark places) [show] a "marked inclination" toward "locative prepositions"– "where," "at," "towards."[111]

Regarding things like Google Books' Ngram viewer, Moretti says "It's like the invention of the telescope… All of a sudden, an enormous amount of matter becomes visible.[112] As far as the implications of all this, one commentator sums things up this way: "…Culturonomics is clearly a discipline with a future, albeit one that hard to fathom for the time being."[113]

Don't you feel free? What could possibly go wrong with this?

A net constricting the humanities?

Obviously, not everyone in the humanities is as excited about this. As Geoffrey Nunberg says, mindful of Seneca's admonition that "too many books spoil the prof", they are "apprehensive about the prospect of turning literary scholarship into an engineering problem". He also comments: "It's hard to imagine anything likelier to raise the hackles of humanists or cultural historians, who aren't disposed to think of their fields on the model, say, of pre-Mendelian biology."[114]

Google seeks to allay the fears of humanities scholars on its blog:

> "We know nothing can replace the balance of art and science that is the qualitative cornerstone of research in the humanities. But we hope the Google Books Ngram Viewer will spark some new hypotheses ripe for in-depth investigation, and invite casual exploration at the same time. We've started working with some researchers already via our Digital Humanities Research Awards, and look forward to additional collaboration with like-minded researchers in the future."[115]

Many in the digital humanities movement seem aware of the skepticism and are eager to help their colleagues understand.   Richard White, a Stanford history professor, although excited about now doing "things we could never do before", says ""We have to overcome a pretty skeptical audience among other historians. We have to be modest in our claims."[116]

Still, even if statements like this might cause some to breathe easier, others still have these concerns about the influence of "Big Data":

- With financial stress and waning student interest, will the "lure of money and technology…. Increasingly push computation front and center"?[117]
- "Will [it] come at the expense of traditional approaches" and "sweep the deck of all money for humanities everywhere else"?[118]

- If things like the Ngram viewer are "the gateway drug that leads to more-serious involvement in quantitative research"[119] will humanities scholars give appropriate attention to their traditional way of working?
- Will scholars form "such a close relationship that the tools" that they "only work with Google-supplied data sets", getting locked-in?[120]
- Even if the "first generation" original thinkers like Moretti show some promise, what about "'dullard' descendants [who] take up 'distant reading' for their research?"[121]

The deeper underlying concern here is probably well-summed up by Geoffrey Nunberg:

> "Data-diggers are gunning to debunk old claims based on 'anecdotal' evidence and answer once-impossible questions about the evolution of ideas, language, and culture. Critics, meanwhile, worry that these stat-happy quants take the human out of the humanities. Novels aren't commodities like bags of flour, they warn. Cranking words from deeply specific texts like grist through a mill is a recipe for lousy research, they say—and a potential disaster for the profession." [122]

So, for all the nuance that men like him seem capable of (see footnote 115), when Lieberman Aiden says: "You can read a small number of books very carefully. Or you can read lots of books "very, very not-carefully", he is going to cause many scholars to feel some concern.[123] It is true as Marc Parry points out, the Science paper "clarifies the limits of quantitative corpus investigations as well as the power" (prompting Parry to say "we're going we'll still need readers"[124]), but in their highly entertaining TED presentation *What we learned from 5 million books* Lieberman Aiden and Michel provide no nuance or caveats at all in their talk about these limits (and about any other concerns persons have with this kind of work). I definitely think that it is possible for Liebermann Aiden's question "Are you willing to examine this data"?[125] to be embraced by humanities scholars, but if they sense the walls are closing in, that question may be received very differently.

Privacy Concerns

There are more concerns to note here. The same kind of data-mining that is used in the Ngram viewer can also be used to produce advertising portfolios on those who read. Concerns of course are only exacerbated by Eric Schmidt's infamous statement that "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."[126]

Recently, Google has united all of their privacy policies into one (which I'm sure all of you Google users took the time to read, right?). Here is how one commentator sums up their reach now:

> "They know what you do online (Google Search), who you correspond with (Google Voice, Gmail, Google Plus), where you go (Google Maps), and what you do (Google Calendar). With the privacy policy change, Google will be using data-

mining algorithms to combine these sources of personal information to create detailed profiles of their users."[127]

Library values as the alternative: humanities with real reading, total quality control, and more

First the humanities.

It seems to me that here we simply have the battle that has been raging forever between the sciences and the humanities (famously articulated by C.P. Snow) taken to the next level. If "culturonomics" gains more and more of a foothold, on what basis will agreements and disagreements in the humanities increasingly be evaluated? *Will they primarily be evaluated on the basis of who has the better algorithmic methods and scientific methodologies? Or will they primarily be evaluated on the basis of the human interpretation that is the result of many hours of study via real reading?*[128]

Shakespeare, John Brockman, and C.P. Snow – look them up! (image: http://gloriamundi.blogsome.com/category/science-artciencia-arte/page/3/, Creative Commons License)

In other words, are scholars like Moretti absolutely determined, *a priori*, to "extract a rational picture" (using "statistics, network theory, computational linguistics and evolution, for example) from what they think they are able to discover via their algorithms?[129] If this is the case, are the chances not good, given that they have already decided to do this, that they *will* extract a rational picture from this chaos and confusion - even if that picture "makes sense" but is not really an accurate picture of reality? (and one that could perhaps be seriously questioned by simply reading just a few of the books that are "mined"?)

I don't think it is unreasonable to have concerns about these things. If people are absolutely convinced that it makes sense to talk about ideas having "an internal genetic code", and that

"metapatterns" are going are able to be discerned everywhere – and that the algorithms that they are able to construct can "net" these patterns – it seems to me that they will at the very least be strongly tempted to believe they can bypass a lot of real reading as well as serious reflection on this reading.[130] The prejudices of all people must give way to science…

Fortunately, at least Moretti does not seem to be the kind of person vulnerable to this temptation. He notes that their early explorations of the "cellars of culture" have not been terribly enlightening, only revealing "a much more *confused* literary landscape than the one we see when we focus only on the canon".[131] Marc Perry writes that "Mr. Moretti is humble enough to admit those "cellars of culture" [that he is exploring] could contain nothing but duller, blander, stupider examples of what we already know. He throws up his hands. 'It's an interesting moment of truth for me,' he says." [132]

It seems a bit much to me. If most every ten word phrase on the internet is absolutely unique, as Walt Crawford convincingly argues[133], I think that should make us think twice about our abilities to not only identify hard-and-fast patterns ("laws") regarding such things , but also to develop effective methods to "net" them. In other words, when it comes to true knowledge and understanding, there is lots of room for the classical humanities where we need to depend on rather large amounts of real reading (or listening) – whether or not everyone discerns that.

Geoffrey Nunberg seems to put things in perspective when he says,

> "I have a friend, a gifted amateur musician and computer scientist, who was involved in electronic music in its early days. Inevitably, within a few years, the field was taken over by composers. That happened partly because new interfaces made the technology more accessible, but also because a command of the subject matter always trumps mere technical expertise. As my friend put it, "It's a lot easier to turn an artist into a geek than to turn a geek into an artist…

> "It's unlikely that 'the whole field' of literary studies—or any other field—will take up these methods, though the data will probably figure in the literature the way observations about origins and etymology do now. But I think Trumpener is quite right to predict that second-rate scholars will use the Google Books corpus to churn out gigabytes of uninformative graphs and insignificant conclusions. But it isn't as if those scholars would be doing more valuable work if they were approaching literature from some other point of view."[134]

I think Nunberg has much to offer here, but I question that last sentence. Is it not possible that spending *more time* doing work from the other "points of view" – namely, those views of the classical humanities – may indeed help them to grow intellectually, and become better scholars who produce more valuable work? In this technology-saturated world, it is precisely the difficulty with "turning geeks into artists" that I think we should be concerned about. In sum, I simply hope Clifford A. Lynch, director of the Coalition for Networked information, is right

when he says "what you may very well see is that this becomes a more commonly accepted tool but not necessarily the center of the work of many people"[135] – but I have my doubts.

In any case, as regards this "war" between science and the humanities, I feel compelled to repeat the quote from E.F. Schumacher:

> "Justice is a denial of mercy, and mercy is a denial of justice. Only a higher force can reconcile these opposites: wisdom. The problem cannot be solved, but wisdom can transcend it. Similarly, societies need stability *and* change, tradition *and* innovation, public interest *and* private interest, planning *and* laissez-faire, order *and* freedom, growth *and* decay. Everywhere society's health depends on the simultaneous pursuit of mutually opposed activities or aims. The adoption of a final solution means a kind of death sentence for man's humanity and spells either cruelty or dissolution, generally both… Divergent problems offend the logical mind."[136]

And it's good to end with a reminder that with Google Books you can indeed just read the books.

Second, the *tools* of the humanities.

As Marc Parry points out: "Google Books… wasn't really built for research. It was built to create more content to sell ads against. And it was built thinking that people would read one book at a time." [137]

Ronald G Musto is rather blunt in his assessment of the matter:

> "the point is that Google Books has represented to us that its massive digitization project will offer a valuable, reliable, open-access research tool that would make the digital at least the equivalent and -- through its ubiquity and ease -- the clear superior of print. It is, after all, the 'public good,' not the 'public good enough,' that lies behind all of Google Books' claims for fair-use rights to its digitization schemes."

He continues:

> "…..Within the scholarly and nonprofit realm over the past decade, there have been dozens of digitization projects: some small, some massive, some open-access, some offered by subscription, some successful, more not so. But several things have united them all: a common purpose for the true good of the community, the highest standards of quality in both technology and content, and a deep-seated and long-abiding concern for the curation, and wide dissemination, of our cultural heritage as a living process that goes beyond commodification."[138]

A couple of these digitization projects Musto speaks of are the Text Creation Partnership (TCP) centered in the University of Michigan and the Corpus of Historical American English (funded by the US National Endowment of the Humanities) at BYU. Shawn Martin discusses how the TCP used the OCR-scans of English books published between 1479 and 1800 from the original microfilm provided by several commercially published databases. They manually transcribed the texts before adding "structural tagging" that would allow the computer "to see elements of the book such as paragraphs, typeface changes, and chapters" –things OCR won't allow. Without good metadata, scholars are not able to search in particular parts of the text like introductions, summaries, quotations, etc. Additionally, OCR also will not allow for proximity searching, nor for scholars to pinpoint every term they search for, nor for the detection of nonstandard typefaces, some foreign languages, or even italics. [139]

Regarding the Corpus of Historical American English, it covers the years from1810-the 2000s[140], and it also "allows for many types of searches that Google Books can't do": wildcard searches, grammatical changes (due to "tags" for parts of speech), and the integration of "synonyms and customized word lists into queries".[141] In sum, as Martin puts it, the question is when it is appropriate to use the Google [Books] product for text searching [and] when it is appropriate to use other products for searching". He has suggestions about how Google Books could be improved using processes similar to those of the TCP[142], but given that it would involve the rather intensive efforts of many, many human beings (to give personal attention to the millions of books Google has digitized) it is unlikely that Google would be interested in this.

Third, the issue of privacy.

The world cares less and less about these kinds of things. Although I did not have time to research it, I think it is likely that companies selling e-Books are tracking data about their readers as well, and again, few probably care.[143] Facebook is the worst offender here, and evidently, almost the whole world is on Facebook. Nevertheless, might not the commitment that librarians have to user privacy[144] be a "selling point" we should tout – especially as some people grow increasingly concerned about such things? Currently, Google's new policy notes that it does not collect user data from Google Books to combine with other services, but it is difficult to see why this seemingly arbitrary decision will stand.[145]

Conclusion

Maybe when you saw the title of this presentation, you thought I might be providing a "How to" list of things to do for students. Well, I hope that you are able to do that with this presentation - I think that you should be able to. But for the most part, I myself have been trying to teach about what I think is important using Google Books. So, not so much teaching about teaching, but just teaching…

I suggest that however we think of education, we can't escape the fact that each one of us are not only guides, but also content providers – we all try to "push our own content" (some more

vigorously and even more offensively than others) even as we, I hope, want to simultaneously encourage, not discourage, critical thinking. So I make no apologies for my pontificating here. I suggest it is what makes us human beings and not machines. If a person finds Google Books interesting, it makes sense to me that they will find the kinds of reflections here interesting. I hope you did.

Google offers us so much, but it does us well to remember what Google's goals are. As Siva Vaidhyanathan reminds us, we are not actually consumers when it comes to Google (those would be its advertisers), but Google's product. Our interests and attention are what Google utilizes and ultimately sells. In addition to using Google for all that it is worth, we may also want to redirect our interests to some of the others sources I've mentioned – and to see their value as well.

Oh, but actually, we may indeed be consumers for Google now since they have been selling eBooks since December 2010 and rumor has it, will be releasing their own tablet as well. Which should not surprise. Librarians know that books, reading, and education through reading are valuable.

I'd like to close with some thought-provoking words from Santosh Desai:

> "An idea like Google Books represents both all that is wonderful and all that is terrifying about the digital revolution…. A knowledge society needs its information in a fluid, readily accessible and easily navigable form. It also needs diversity, freedom and the chaotic cadence of a million voices that sing their own determined tunes. The question before us is not an easy one. Either way, we will all win and we will all lose."[146]

It is always a double-edged sword with Google. Hopefully, as we look towards the future, we will continue thinking hard about the best way to work with all the changes that have come our way – and to act as responsibly as we can.

Thank you.

Select Bibliography (more citations found in endnotes)

Bivens-Tatum, Wayne. "Libraries and the Commodification of Culture, Academic Librarian (blog), February 13, 2012, http://blogs.princeton.edu/librarian/2012/02/libraries-and-the-commodification-of-culture/

Coyle, Karen. "Google Files Motion to Dismiss." *Coyle's InFormation* (blog), December 26, 2011 (2:16 p.m.), http://kcoyle.blogspot.com/2011/12/google-files-motion-to-dismiss.html

Desai, Santosh. "Column: Are Books on Google Good for Us?" *Financial Express*, (Feb 02, 2010) n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/872800719?accountid=26720.

Darnton, Robert. "A Library Without Walls," *NYR Blog* (blog), October 4, 2010 (9:20 a.m.), http://www.nybooks.com/blogs/nyrblog/2010/oct/04/library-without-walls/.

Darnton, Robert. "Can We Create a National Digital Library?" *New York Review of Books,* October 28, 2010, http://www.nybooks.com/articles/archives/2010/oct/28/can-we-create-national-digital-library/

Darnton, Robert. "Six Reasons Google Books Failed." *NYR Blog* (blog), March 28, 2011, (11:00 a.m.), http://www.nybooks.com/blogs/nyrblog/2011/mar/28/six-reasons-google-books-failed/

Darnton, Robert. "The Library: Three Jeremiads." *New York Review of Books.* 57, no. 20 (November 2010): pp. 22-27.

Efrati, Amir. "Judge Rejects Google Books Settlement." *Wall Street Journal*, Mar. 23, 2011. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/858106644?accountid=26720

"Google Book Privacy Still a Concern Post GBS," *Open Book Alliance* (blog), October 27, 2011. http://www.openbookalliance.org/2011/04/google-book-privacy-still-a-concern-post-gbs/

 "Google's Big Book Case." Economist, September 3, 2009. http://www.economist.com/node/14363287

Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

Haven, Cynthia, "Non-consumptive Research? Text Mining? Welcome to the Hotspot of Humanities Research at Stanford*" Stanford University News*, December 1, 2010. http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html

Howard, Jennifeer. "With No Google Books Deal, Libraries Push New Plans for Digital Access." *Chronicle Of Higher Education* 57, no. 30 (April 2011): A12. *Academic Search Premier*, EBSCO*host* (accessed March 7, 2012)

Jackson, Millie. "Using Metadata to Discover the Buried Treasure in Google Book Search". *Journal of Library Administration.* 47, no. ½ (2008): 165-173.

Kloha, Jeff, "Words, Words, Words", *Concordia Theology* (blog), December 21, 2010, (6:00 a.m.), http://concordiatheology.org/2010/12/words-words-words/

Lessig, Lawrence. "For the Love of Culture." *The New Republic* , January 26, 2010 (12:00 am) http://www.tnr.com/print/article/the-love-culture

Martin, Shawn. "To Google or Not to Google, That is the Question: Supplementing Google Book Search to Make it More Useful for Scholarship." *Journal of Library Administration* 47, no 1-2 (2008): 141-150

Mattson, Kevin. "Paying the Piper: Is Culture Ever Free?." *Dissent (00123846)* 58, no. 2 (Spring 2011): 69-73. *Academic Search Premier*, EBSCO*host* (accessed March 9, 2012).

Meyer, Frankie. "Use Google as a resource for hard-to-find books" The Joplin Globe, February 27, 2012, http://www.joplinglobe.com/lifestyles/x2118802287/Frankie-Meyer-Use-Google-as-a-resource-for-hard-to-find-books

Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

Nunberg, Geoffrey. "Google Book Search: A Disaster for Scholars." *Chronicle of Higher Education*, August 31, 2009. http://chronicle.com/article/Googles-Book-Search-A/48245/

Parry, Marc. "Google Starts Grant Program for Studies of Its Digitized Books." *Chronicle of Higher Education*, March 31, 2010. http://chronicle.com/article/Google-Starts-Grant-Program/64891/

Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly*. 49, no. 1 (2008): 1-13.

Schumacher, E. F. *A Guide for the Perplexed*. New York: Harper & Row, 1977,127.

Singer, Natasha. "Playing Catch-Up in a Digital Library Race." New York Times, Jan. 8, 2011.

Swift, Mike. 2010. "Google Books may Advance Humanities Research, Scholars Say." *McClatchy - Tribune News Service*, Aug 05, n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/737529948?accountid=26720

Thompson, Chris. "The Case Against Google Books; How three East Bay librarians led the revolt against the company's plans to archive all earthly knowledge." *East Bay Express* (California). October 14: *LexisNexis Academic. Web.* (accessed March 7, 2012).

"Tome Raider." *Economist*, September 3, 2009. http://www.economist.com/node/14376406

Wile, Rob, "Google Books Reveals How Words Have Changed in Popularity Over Time", Business Insider (blog), January 25, 2012, (8:19 a.m.), http://www.businessinsider.com/charts-google-books-reveals-the-most-popular-words-in-history-2012-1

[1] Singer, Natasha. "Playing Catch-Up in a Digital Library Race." New York Times, Jan. 8, 2011. http://www.nytimes.com/2011/01/09/business/09stream.html

[2] Darnton, Robert. "A Library Without Walls," *NYR Blog* (blog), October 4, 2010 (9:20 a.m.), http://www.nybooks.com/blogs/nyrblog/2010/oct/04/library-without-walls/.

[3] *The New Dictionary of Cultural Literacy, Houghton Mifflin*. Boston: Houghton Mifflin, 2002. s.v. "liberal arts," http://www.credoreference.com/entry/hmndcl/liberal_arts (accessed March 02, 2012).

[4] Lessig, Lawrence. "For the Love of Culture." *The New Republic* , January 26, 2010 (12:00 am) http://www.tnr.com/print/article/the-love-culture

[5] Thompson, Chris. "The Case Against Google Books; How three East Bay librarians led the revolt against the company's plans to archive all earthly knowledge." *East Bay Express* (California). (October 14, 2009 Wednesday): *LexisNexis Academic. Web.* (accessed March 7, 2012).

[6] Brin, Sergey. "A Library to Last Forever." *International Herald Tribune* (Paris), Oct 10, 2009: 8.

[7] Beck, Richard. "A bookshelf the Size of the World." The Boston Globe (Boston) , July 24, 2011. http://articles.boston.com/2011-07-24/bostonglobe/29810463_1_google-books-robert-darnton-digitization

[8] Nunberg, Geoffrey. "Google Book Search: A Disaster for Scholars." *Chronicle of Higher Education*, August 31, 2009. http://chronicle.com/article/Googles-Book-Search-A/48245/

[9] Howard, Jennifeer. "With No Google Books Deal, Libraries Push New Plans for Digital Access." *Chronicle Of Higher Education* 57, no. 30 (April 2011): A12. *Academic Search Premier*, EBSCO*host* (accessed March 7, 2012)

[10] Rocco Staino, et al. "Libraries a side issue at Google Settlement Hearing." *Library Journal* 135, no. 5 (March 15, 2010): 12. *Academic Search Premier*, EBSCO*host* (accessed March 7, 2012)

[11] As suggested here: Vaidhyanathan, Siva. "Thank You, Judge Chin: Why the Google Books Setback is an Opportunity to Reassert Academic Values." *Chronicle of Higher Education,* March 24, 2011. http://chronicle.com/article/Thank-You-Judge-Chin/126888/

[12] Suber, Peter, "Michigan President Defends Google Library to AAP," *Open Access News: News from the Open Access Movement (*blog), February 7, 2006 (8:46 a.m.), http://www.earlham.edu/~peters/fos/2006/02/michigan-president-defends-google.html, found originally here: Crawford, Walt. "Discovering Books: The OCA/GBS Saga Continues." *Cites & Insights: Crawford at Large* 6, no. 6 (Spring 2006). http://citesandinsights.info/v6i6a.htm)

[13] Quint, Barbara. "An Unapologetic Apologia for Google Book Search." *Information Today* 26, no. 9 (October 2009): 7-8. *Academic Search Premier*, EBSCO*host* (accessed March 9, 2012)

[14] This You Tube video, where the same search is performed, is a nice tour, even if some of the layout of the Google Books page has changed (along with their ever-evolving search algorithms): http://www.youtube.com/watch?v=GJS7-MR4tNk

[15] Quint, Barbara. "An Unapologetic Apologia for Google Book Search." *Information Today* 26, no. 9 (October 2009): 7-8. *Academic Search Premier*, EBSCO*host* (accessed March 9, 2012)

[16] Efrati, Amir and Jeffrey A. Trachtenberg. "Judge Rejects Google Books Settlement." *Wall Street Journal,* Mar 23, 2011. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/858106644?accountid=26720

[17] Kolowich, Steve. "Please Refine Your Search Terms." *Inside Higher Ed*, March 23, 2011. http://www.insidehighered.com/news/2011/03/23/judge_rejects_google_books_settlement

[18] For more info, see here: "Judge Chin's Ruling By the Number," *Open Book Alliance* (blog), March 24, 2011. http://www.openbookalliance.org/2011/03/judge-chins-ruling-by-the-numbers/

[19] This PowerPoint presentation does a nice job of summing up the primary different views regarding the settlement: Jones, Elisabeth. *The Google Books Library Project: Personal Experiences, Public Perspectives, and Issues Relevant to Libraries*. Slideshare presentation. Winter 2010. .http://www.slideshare.net/elisabethjones/google-books-lecture. The following is how the Open Book Alliance (made up of those opposed to the Google settlement, and even including businesses as well such as Yahoo and Microsoft) summed up the situation: "Since the beginning of this process, criticism of the Google Book Settlement fell roughly into three categories – objectors who viewed the terms of the deal as a complete abrogation of existing copyright law, those who focused on the role that Google Book Search and the settlement terms played in Google's continuing abuse of its market dominance in violation of antitrust laws, and those who viewed the settlement as a total misuse of the class action mechanism." "Final KO for GBS?" *Open Book Alliance* (blog), September 12, 2011. http://www.openbookalliance.org/2011/09/final-ko-for-gbs/

[20] Band, Jonathan. "A Guide for the Perplexed: Libraries and the Google Library Project Settlement." *LLRX.com: Law and Technology Resources for Legal Professionals*, December 14, 2008. http://www.llrx.com/features/googleprojectsettlement.htm.

[21] All of the points mentioned above were brought up in this article: Darnton, Robert. "Six Reasons Google Books Failed." *NYR Blog* (blog), March 28, 2011, (11:00 a.m.), http://www.nybooks.com/blogs/nyrblog/2011/mar/28/six-reasons-google-books-failed/

[22] Kolowich, Steve. "Please Refine Your Search Terms." *Inside Higher Ed*, March 23, 2011. http://www.insidehighered.com/news/2011/03/23/judge_rejects_google_books_settlement

[23] "Google's Big Book Case: the Internet Giant's Plan to Create a Vast Digital Library Should be Given a Green Light." *The Economist*, September 3, 2009. http://www.economist.com/node/14363287

[24] Darnton, Robert. "Six Reasons Google Books Failed." *NYR Blog* (blog), March 28, 2011, (11:00 a.m.), http://www.nybooks.com/blogs/nyrblog/2011/mar/28/six-reasons-google-books-failed/

[25] "The two plaintiffs were recognized as the representatives of a class consisting of nothing less than every single copyright holder in the universe. As bizarre as it sounds, every author, publisher, or copyright holder on Earth was suddenly having its rights negotiated in court, even if most of them didn't know it or were even aware that a lawsuit was underway. With this settlement, Google got the right to sell access to tens of millions of orphan books and academic studies in one fell swoop, even if the authors didn't know it. Google spokeswoman Jennie Johnson said nonprofits like the Internet Archive are free to go through the same negotiation process. "Anybody can do exactly what we did," Johnson said. But according to Brantley, that's essentially impossible. "Because of the nature of the settlement as a class action, Google would have a unique release of liability from the commercialization of works that went unclaimed in the process of notification of authors and publishers," Brantley said. " (Chris Thompson, "The Case Against Google Books; How three East Bay librarians led the revolt against the company's plans to archive all earthly knowledge.." East Bay Express (California). (October 14, 2009 Wednesday): 4399 words. LexisNexis Academic. Web. Date Accessed: 2012/03/07). Also see the view of the Open Book Alliance: http://www.openbookalliance.org/2011/02/fox-in-the-henhouse-authors-guild-to-testify-on-ip-rights/ A quote: "…To be frank, the Authors Guild sold its member authors down the river. As followers of the GBS are well aware, the Authors Guild's negotiators (who, as NWU, ASJA, and SFFWA

note "represent only a tiny fraction of published writers") worked with Google to craft a settlement that locks authors into the Books Rights Registry – unless, of course, these authors find out about the settlement and navigate the GBS' incomprehensible website. In fact, the process is so impenetrable that Google and the Authors Guild had to file for an extension last week to extend the deadline for an "up front" cash payment for unauthorized digitization of entire works.

In short, by partnering with Google in the GBS, the Authors Guild has damaged its credibility when it comes to speaking for rightsholders on intellectual property issues…"

[26] All six of these points are mentioned in the following article: Darnton, Robert. "Six Reasons Google Books Failed." *NYR Blog* (blog), March 28, 2011, (11:00 a.m.), http://www.nybooks.com/blogs/nyrblog/2011/mar/28/six-reasons-google-books-failed/

[27] Meola, Marc. "Siva Vaidhyanathan Questions Google Book Search." ACRLog (blog), April 23, 2007. http://acrlog.org/2007/04/23/siva-vaidhyanathan-questions-google-book-search/

[28] Smith, Charles Edward. "A Few Thoughts on the Google Library Project." EDUCASE Quarterly, 31, no. 1 (January – March 2008): pp. 10—11. http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/AFewThoughtsontheGoogleBooksLi/162510

[29] Efrati, Amir and Jeffrey A. Trachtenberg. " Judge Rejects Google Books Settlement." *Wall Street Journal,* Mar 23, 2011. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/858106644?accountid=26720

[30] "Tome Raider." *Economist*, September 3. 2009. http://www.economist.com/node/14376406 Another comment from another Economist article: "most of the 'orphan books' were unlikely to be of much value. If any are, their appearance in Google's archives is likely to bring long-lost copyright owners out of the woodwork to claim the proceeds. They can then sell the digital rights to their books to firms other than Google, increasing competition. " "Google's Big Book Case." Economist, September 3, 2009. http://www.economist.com/node/14363287

[31] "Joseph Esposito, a consultant who advises scholarly publishers", says ""I think that is just plain wrong. There is a reason books went out of print in the past, a reason that orphans are orphans. These are books of marginal value to higher education, as the historical lack of demand demonstrates." http://www.insidehighered.com/news/2011/03/23/judge_rejects_google_books_settlement

[32] Kolowich, Steve. "Please Refine Your Search Terms." *Inside Higher Ed*, March 23, 2011. http://www.insidehighered.com/news/2011/03/23/judge_rejects_google_books_settlement

[33] Efrati, Amir. "Judge Rejects Google Books Settlement." *Wall Street Journal*, Mar. 23, 2011. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/858106644?accountid=26720

[34] Who may now be more aggressive, given the judge's ruling. Kolowich, Steve. "Please Refine Your Search Terms." *Inside Higher Ed*, March 23, 2011. http://www.insidehighered.com/news/2011/03/23/judge_rejects_google_books_settlement

[35] France-Presse, Agence. "U.S. Universities Hit with Copyright Infringement Suit", The Raw Story (blog), September 12, 2011, (8:44 p.m.) http://www.rawstory.com/rs/2011/09/12/u-s-universities-hit-with-copyright-infringement-suit/

[36] Coyle, Karen. "Google Files Motion to Dismiss." *Coyle's InFormation* (blog), December 26, 2011 (2:16 p.m.), http://kcoyle.blogspot.com/2011/12/google-files-motion-to-dismiss.html

[37] Kenall, Brent, and Bravin, Jess. "Foreign Copyrights Upheld: Court Backs Reciprocal International Protection for Previously Public Material." *Wall Street Journal*, Jan. 19, 2012. http://online.wsj.com/article/SB10001424052970204555904577168752017626174.html ; Ritti, Melissa. "Supreme Court Sides With Government in Dispute Over Public Domain Works." *LexisNexis Communities: Litigation Resource Community* (blog), Jan. 18, 2012 (8:25 p.m.), http://www.lexisnexis.com/community/litigationresourcecenter/blogs/litigationblog/archive/2012/01/18/supreme-court-sides-with-government-in-dispute-over-public-domain-works.aspx ; see also Crawford, Walt. "Talking About the Public Domain." *Cites & Insights: Crawford at Large* 11, no. 7 (August 2011). http://citesandinsights.info/v11i7b.htm

[38] Again, the view of the Open Book Alliance: "Law professor and copyright expert Pamela Samuelson, who we've written about here before, recently participated in webinar moderated by Publishers Weekly on the rejection of the Google Book Settlement. During the panel, she called on the parties involved to reach a new settlement and, more importantly, for the US Copyright Office to get involved so that fundamental copyright reforms can be had through legislation, not litigation. Noting that Judge Chin indicated a preference for a new settlement premised on opt-in, she went on to say that "the [U.S.] Copyright Office is the best entity to find out what the parties want and what may be a good model for legislation." This jibes with what Judge Chin wrote in his opinion rejecting the GBS when he said, "… the establishment of a mechanism for exploiting unclaimed books is a **matter more suited for Congress than this Court**. . . . The questions of who should be entrusted with guardianship over orphan books, under what terms, and with what safeguards are matters **more appropriately decided by Congress than through an agreement among private, self-interested parties**." We've long argued (as has Prof. Samuelson) that the GBS represented a mechanism to sidestep Congressional authority over copyright reform, and we rejoin her call for orphan book legislation now that the flawed GBS has been rejected." http://www.openbookalliance.org/2011/05/prof-samuelson-renews-call-for-copyright-legislation/

[39] Desai, Santosh. "Column: Are Books on Google Good for Us?" *Financial Express*, (Feb 02, 2010) n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/872800719?accountid=26720.

[40] Thompson, Chris. "The Case Against Google Books; How three East Bay librarians led the revolt against the company's plans to archive all earthly knowledge." *East Bay Express* (California). October 14: *LexisNexis Academic. Web.* (accessed March 7, 2012).

[41] see Lessig, Lawrence. "For the Love of Culture." *The New Republic* , January 26, 2010 (12:00 am) http://www.tnr.com/print/article/the-love-culture

[42] Darnton, Robert. "The Library: Three Jeremiads." *New York Review of Books.* 57, no. 20: pp. 22-27. Elsewhere he was more blunt: "a company intent on maximizing profit at the expense of the public" Darnton, Robert. "Six Reasons Google Books Failed." *NYR Blog* (blog), March 28, 2011, (11:00 a.m.), http://www.nybooks.com/blogs/nyrblog/2011/mar/28/six-reasons-google-books-failed/. Again, the Open Book Alliance: "Today, as the Open Book Alliance sees it: every time that Google moves into another

vertical market (like Book Search) it's really a new source of consumer information that it can monetize and use to strengthen its dominant position. Seven years removed, the cross section of consumer advocates, librarians, authors and publishers that have publicly opposed the settlement is indicative of how far the Google Book Search project has fallen."
http://www.openbookalliance.org/2011/09/adventures-in-google%E2%80%99s-audacity/

[43] Lessig, Lawrence. "For the Love of Culture." *The New Republic* , January 26, 2010 (12:00 am)
http://www.tnr.com/print/article/the-love-culture

[44] Quoted in Darnton, Robert. "Can We Create a National Digital Library?" *New York Review of Books,* October 28, 2010, http://www.nybooks.com/articles/archives/2010/oct/28/can-we-create-national-digital-library/

[45] This has deeper implications as well: I think it is a safe bet that most people not only do not want "the monetization of reading transactions", but also do not want the "commodification of culture" where "every human exchange a market transaction". In short, they do not really desire this "increasing commodification of human relationships". Mentioned by Jeremy Rifkin and Wayne Bivens-Tatum here: Bivens-Tatum, Wayne. "Libraries and the Commodification of Culture, Academic Librarian (blog), February 13, 2012, http://blogs.princeton.edu/librarian/2012/02/libraries-and-the-commodification-of-culture/

[46] Darnton, Robert. "The Library: Three Jeremiads." *New York Review of Books.* 57, no. 20 (November 2010): pp. 22-27.

[47] See a defense of the Google project as fair use here: Crawford, Walt. "Discovering Books: The OCA/GBS Saga Continues." *Cites & Insights: Crawford at Large* 6, no. 6 (Spring 2006). http://citesandinsights.info/v6i6a.htm)

[48] Lessig, Lawrence. "For the Love of Culture." *The New Republic* , January 26, 2010 (12:00 am)
http://www.tnr.com/print/article/the-love-culture

[49] Whitebloom, Kenny. "Press: 'Nothing Like it Has Ever Existed," *DPLA* (blog), January 18, 2012, http://dp.la/2012/01/18/press-nothing-like-it-has-ever-existed/

[50] Full quote: "Creativity requires stability. You can't express yourself— write the book or article or teach the class—if you constantly worry about the next source of income. Creativity is not the same thing as the hyperactive search for the next gig that can get you through the short term. It's about the expansive freedom that stability provides. It's about having time to reflect and think.   Mattson, Kevin. "Paying the Piper: Is Culture Ever Free?." *Dissent (00123846)* 58, no. 2 (Spring 2011): 69-73. *Academic Search Premier*, EBSCO*host* (accessed March 9, 2012).

[51] Mangalwadi, Vishal. *Truth and transformation: a manifesto for ailing nations*. Seattle, Wash: YWAM Pub, 2009.

[52] Lessig, Lawrence. "For the Love of Culture." *The New Republic* , January 26, 2010 (12:00 am)
http://www.tnr.com/print/article/the-love-culture

[53] Pogge, Thomas, and Saunders, Alan. *The Right to Property and the Right to Health*. Podcast audio. Philosopher's Zone. MP3, February 6, 2010. Accessed March 12, 2012. http://www.abc.net.au/radionational/programs/philosopherszone/the-right-to-property-and-the-right-to-health/3103122.

[54] Fowler, Geoffrey A. "Libraries Have a Novel Idea," *Wall Street Journal*, June 29, 2010.

http://online.wsj.com/article/SB10001424052748703279704575335193054884632.html ; see also "Internet Archive and Library Partners Develop Joint Collection of 80,000+ eBooks to Extend Traditional In-Library Lending Model," *Internet Archive* (blog), February 22, 2011, (4:16 p.m.), http://www.archive.org/post/349420/in-library-ebook-lending-program-launched .

[55] It seems to me that the publishers and authors both deserve their wages – even if publishers today may not take as many risks as those in the past, the fact of the matter is that if they simply do their job well, regardless of the amount of risk, they should be paid the wage they deserve.

[56] Schumacher, E. F. *A Guide for the Perplexed*. New York: Harper & Row, 1977, 127. Also note this: "As David Bollier notes in *Silent Theft*, "the issue is not market versus commons. The issue is how to set equitable and appropriate boundaries between the two realms—semi-permeable membranes—so that the market and the commons can each retain integrity while invigorating that other. That equilibrium is now out of balance as businesses try to exploit all available resources, including those that everyone owns and uses in common." Bivens-Tatum, Wayne. "Libraries and the Commodification of Culture, Academic Librarian (blog), February 13, 2012, http://blogs.princeton.edu/librarian/2012/02/libraries-and-the-commodification-of-culture/

[57] Singer, Natasha. "Playing Catch-Up in a Digital Library Race." New York Times, Jan. 8, 2011

[58] Darnton, Robert. "The Library: Three Jeremiads." *New York Review of Books.* 57, no. 20 (November 2010): pp. 22-27.

[59] Which again, is not uncontroversial itself – the Open Book Alliance, which includes Microsoft Corp., Amazon.com Inc. and Yahoo Inc., will say this, in addition to some publishers and authors.

[60] Efrati, Amir. "Judge Rejects Google Books Settlement." *Wall Street Journal*, Mar. 23, 2011. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/858106644?accountid=26720

[61] Marlowe, L. "Washington diary." *Irish Times,* Feb 26, 2011. P 18. Retrieved from http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/853784373?accountid=26720

[62] Madrigal, Alexis. "Inside the Google Books Algorithm." *The Atlantic* (blog), November 1, 2010, (3:00 p.m.), http://www.theatlantic.com/technology/archive/2010/11/inside-the-google-books-algorithm/65422/

[63] Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly.* 49, no. 1 (2008): 1-13.

[64] "Google books is a great source." *The Ledger,* Jan 30, 2011. pp. n/a. Retrieved from http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/847983062?accountid=26720

[65] Kloha, Jeff, "Words, Words, Words", *Concordia Theology* (blog), December 21, 2010, (6:00 a.m.), http://concordiatheology.org/2010/12/words-words-words/

[66] Bergson-Michelson, Tasha, "Can't Confirm That Quotation? Search Google Books." *Mind/Shift* (blog), March 7, 2012, (5:00 a.m.), http://mindshift.kqed.org/2012/03/cant-confirm-that-quotation-search-google-books/

Here are some fun exercises from that article:

- Uncover the original publication date of the print story, "Yes, Sweetie, Trees Can Have Purple Leaves," by James J. Backen.

- Locate the story "Pat and the Sarpint," which the reader recalls being in a book called Early American Folk Tales.

- Prove that a piece about a Jewish tradition around first menstruation, entitled "The Slap," has actually been published somewhere.

- Identify a book described by the reader in these words: "It contained Native Americans as the primary protagonists, included witchcraft, gun-running and I believe a supporting cast of hippies."

In some of these cases, some fancy searching was involved (stay tuned for more about locating gun-runners, witches, and hippies another day). In most of these cases, however, a very straightforward Google Books search, such as typing the title of the book or story into the basic Google Books search box, uncovered the hoped-for source instantly."

[67] Jackson, Millie. "Using Metadata to Discover the Buried Treasure in Google Book Search". *Journal of Library Administration.* 47, no. ½ (2008): 165-173.

[68] Meyer, Frankie. "Use Google as a resource for hard-to-find books" The Joplin Globe, February 27, 2012, http://www.joplinglobe.com/lifestyles/x2118802287/Frankie-Meyer-Use-Google-as-a-resource-for-hard-to-find-books

[69] Jackson, Millie. "Using Metadata to Discover the Buried Treasure in Google Book Search". *Journal of Library Administration.* 47, no. ½ (2008): 165-173.

[70] Jackson, Millie. "Using Metadata to Discover the Buried Treasure in Google Book Search". *Journal of Library Administration.* 47, no. ½ (2008): 165-173.

[71] Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly.* 49, no. 1 (2008): 1-13.

[72] Nunberg, Geoffrey. "Google Book Search: A Disaster for Scholars." *Chronicle of Higher Education*, August 31, 2009. http://chronicle.com/article/Googles-Book-Search-A/48245/

[73]http://books.google.com/ngrams/graph?content=George+Washington,Thomas+Jefferson,Abraham+Lincoln&year_start=1776&year_end=2008&corpus=0&smoothing=3   found here: "Find Out What's in a

Word, or Five, with the Google Books Ngram Viewer," *Google Official Blog* (blog), December 16, 2010, (1:08 p.m.) http://googleblog.blogspot.com/2010/12/find-out-whats-in-word-or-five-with.html

[73] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[75] Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly.* 49, no. 1 (2008): 1-13.

[76] "…. And while Google's machine classification system will certainly improve, extracting metadata mechanically isn't sufficient for scholarly purposes. After first seeming indifferent, Google decided it did want to acquire the library records for scanned books along with the scans themselves, but as of now the company hasn't licensed them for display or use—hence, presumably, those stabs at automatically recovering publication dates from the scanned texts." Nunberg, Geoffrey. "Google Book Search: A Disaster for Scholars." *Chronicle of Higher Education*, August 31, 2009. http://chronicle.com/article/Googles-Book-Search-A/48245/

[77] Nunberg, Geoffrey. "Google Book Search: A Disaster for Scholars." *Chronicle of Higher Education*, August 31, 2009. http://chronicle.com/article/Googles-Book-Search-A/48245/

[78] Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly.* 49, no. 1 (2008): 1-13.

[79] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[80] Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly.* 49, no. 1 (2008): 1-13.

[81] See this fun site, The Art of Google Books: http://theartofgooglebooks.tumblr.com/

[82] Thompson, Chris. "The Case Against Google Books; How three East Bay librarians led the revolt against the company's plans to archive all earthly knowledge." *East Bay Express* (California). October 14: *LexisNexis Academic. Web.* (accessed March 7, 2012).

[83] Martin, Shawn. "To Google or Not to Google, That is the Question: Supplementing Google Book Search to Make it More Useful for Scholarship." *Journal of Library Administration* 47, no 1-2 (2008): 141-150

[84] Wilkin, John. "Hathi Trust and the Google Deal", *Library Journal*, Jan. 1, 2009.

[85] Read more about this here: Haven, Cynthia, "Stanford Technology Helps Scholars Get 'Big Picture' of the Enlightenment*." Stanford University News*, December 17, 2009. http://news.stanford.edu/news/2009/december14/republic-of-letters-121809.html

[86] Oxford supplied the metadata for about 50,000 letters, allowing the project to go "beyond any of our expectations," said Edelstein. Haven, Cynthia, "Stanford Technology Helps Scholars Get 'Big Picture' of the Enlightenment*." Stanford University News*, December 17, 2009. http://news.stanford.edu/news/2009/december14/republic-of-letters-121809.html

[87] Swift, Mike. 2010. "Google Books may Advance Humanities Research, Scholars Say." *McClatchy - Tribune News Service*, Aug 05, n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/737529948?accountid=26720.

[88] Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[89] Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[90] Cohen, Patricia. "Google Database Puts Language in a Petri Dish." *International Herald Tribune*, Dec 18, 2010: 12. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/818453126?accountid=26720

[91] "the application of high-throughput data collection and analysis to the study of human culture," which "extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[92] "That's culturomics with a long *o*, with the implication that the object of study is the "culturome," presumably the mass of structured information that characterizes a culture. The point of comparison might be biological models of evolution or simply the idea that culture, like the genome, can be "cracked" via massive distributed (that is, "high-throughput") processing." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

 "Yet some humanities researchers in the traditional camp complain that their field can never be encapsulated by the frequency charts of words and phrases produced by the n-grams tool. "I think saying all books equal the DNA of human experience — I think that's a very dangerous parallel," says Cohen. How do you factor in the cultural contributions of furniture, or dance, or ticket stubs at a movie hall, he asks. What about all the books that were never published? Or the culture as experienced by the world's vast illiterate populations?" Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

[93] Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

[94] "Culturonomics and the Google Book Project," The Physics arXiv Blog (blog), February 27, 2012. http://www.technologyreview.com/blog/arxiv/27608/?p1=blogs

[95] Veith, Gene, "Ngrams," *Cranach: the Blog of Veith* (blog), December 28, 2010, http://www.geneveith.com/2010/12/28/ngrams/ See also: http://books.google.com/ngrams/info

[96] Wile, Rob, "Google Books Reveals How Words Have Changed in Popularity Over Time", Business Insider (blog), January 25, 2012, (8:19 a.m.), http://www.businessinsider.com/charts-google-books-reveals-the-most-popular-words-in-history-2012-1

[97] A thoughtful comment from this blog post: Veith, Gene,"Ngrams," *Cranach: the Blog of Veith* (blog), December 28, 2010, http://www.geneveith.com/2010/12/28/ngrams/ : "…try the word "aesthetic" from 1500 to 2000. Was the world unconcerned with beauty before 1850? Does the steep rise in the use of the word after 1850 tell us about a change in the world – or does it only tell us something (what exactly?) about the authors of books after 1850?"

[98] Pope J.T., and Holley R.P. "Google book search and metadata". *Cataloging and Classification Quarterly*. 49, no. 1 (2008): 1-13.

[99] "Find Out What's in a Word, or Five, with the Google Books Ngram Viewer," *Google Official Blog* (blog), December 16, 2010, (1:08 p.m.) http://googleblog.blogspot.com/2010/12/find-out-whats-in-word-or-five-with.html . "Summing up the results in their Science paper, one commented "their approach can produce unique insights into fields such as the evolution of grammar, the adoption of technology, the pursuit of fame, censorship and historical epidemiology." Here: "Culturonomics and the Google Book Project," The Physics arXiv Blog (blog), February 27, 2012. http://www.technologyreview.com/blog/arxiv/27608/?p1=blogs

[100]Also see http://medicalxpress.com/news/2012-01-happy-language-isnt.html

[100] See more on Jocker's work here: Haven, Cynthia, "Non-consumptive Research? Text Mining? Welcome to the Hotspot of Humanities Research at Stanford*" Stanford University News*, December 1, 2010. http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html

[101] Parry, Marc. "Google Starts Grant Program for Studies of Its Digitized Books." *Chronicle of Higher Education*, March 31, 2010. http://chronicle.com/article/Google-Starts-Grant-Program/64891/   Also, Dan Cohen a professor of history, talks about overcoming "this problem of anecdotal history….less based on a chosen few authors…" Swift, Mike. 2010. "Google Books may Advance Humanities Research, Scholars Say." *McClatchy - Tribune News Service*, Aug 05, n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/737529948?accountid=26720.

[102] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/
; Also see: Swift, Mike. 2010. "Google Books may Advance Humanities Research, Scholars Say." *McClatchy - Tribune News Service*, Aug 05, n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/737529948?accountid=26720

[103] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[104] Swift, Mike. 2010. "Google Books may Advance Humanities Research, Scholars Say." *McClatchy - Tribune News Service*, Aug 05, n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/737529948?accountid=26720

[105] See here for detail: Parry, Marc. "Google Starts Grant Program for Studies of Its Digitized Books." *Chronicle of Higher Education*, March 31, 2010. http://chronicle.com/article/Google-Starts-Grant-Program/64891/

[106] Parry, Marc. "Google Starts Grant Program for Studies of Its Digitized Books." *Chronicle of Higher Education*, March 31, 2010. http://chronicle.com/article/Google-Starts-Grant-Program/64891/

[107] Haven, Cynthia, "Non-consumptive Research? Text Mining? Welcome to the Hotspot of Humanities Research at Stanford*" Stanford University News*, December 1, 2010. http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html

[108] continued: "…The patterns that emerged will be grist for theories of language evolution, But quantitative methods are already widely accepted in the field, and in any case, morphological change is a "cultural" phenomenon only by courtesy of the dean of humanities." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[109] Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

[110] Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[111] Haven, Cynthia, "Non-consumptive Research? Text Mining? Welcome to the Hotspot of Humanities Research at Stanford*" Stanford University News*, December 1, 2010. http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html

[112] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[113] "Culturonomics and the Google Book Project," The Physics arXiv Blog (blog), February 27, 2012. http://www.technologyreview.com/blog/arxiv/27608/?p1=blogs

[114] Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[115] Back at the Aiden house, the Shabbat dinner guests have all laved their hands with a glass of water and returned to the sun deck for matzo-ball soup. Lieberman Aiden explains some of the trepidation he felt when he and Michel talked to the historians at Princeton about their work. "I was a little bit nervous going in," he says. "I really thought that we were going to get denounced at one point."

Although Lieberman Aiden and Michel are sensitive to the feelings of traditional humanities scholars, they are also too young, restless and deeply ambitious to slow their own pursuits. Lieberman Aiden says that the influence of technology on the humanities is already past a tipping point. The tools and methods that it provides, he says, will be impossible for researchers to ignore. And yet he doesn't think that the old approaches will ever disappear. "I think you should use the best methods available — and all of them," he says. "And I think that includes carefully reading texts and trying to get behind what authors think."

Daniel Koll, one of the dinner guests, shyly interrupts. "Erez? Do you think you're maybe partly influenced in that kind of thinking by your religious upbringing? From my limited outsider's perspective, Judaism has a very strong interpretive component. There is no single authority on a text, and so on." He wonders whether Lieberman Aiden, like any good humanities scholar, enjoys wrestling with the ambiguities of religious texts as much as he enjoys cool, hard data.

Clearly, the answer is yes — why else would his host have spent a year of his life at Yeshiva University in New York, studying the Talmud and Jewish case law? But Lieberman Aiden, who prefers to talk about other people and their ideas rather than himself, provides an indirect answer by way of history. He tells the story of Isaac Casaubon, a sixteenth-century Protestant scholar, who undermined the presumed Egyptian provenance of a set of religious texts by identifying a reference to a Greek play on words — something that could only have been written hundreds of years later. "That point is as objective an interpretive remark as any remark a scientist might make," says Lieberman Aiden. "So the methods of humanists are very, very formidable. And I think the degree of insecurity they have over whether these methods are here to stay is not really befitting."

…. And yet the boundary between work and play — just like that between the sciences and the humanities — is not one that Lieberman Aiden respects. That might just be what makes him successful, says Lander. For centuries, the best science has come from the most playful scientists, he says. Think of Watson and Crick shirking the lab in favour of tennis; think of Einstein and his wild-haired bike rides.

"What do children do?" says Lander. "They learn, they're curious, they're stimulated. The problem is, at some point, many people get in a rut. They're not really interested in learning more. They're not able to be fascinated and delighted by everything around them. Erez — he hasn't lost the playfulness." Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

[116] Swift, Mike. 2010. "Google Books may Advance Humanities Research, Scholars Say." *McClatchy - Tribune News Service*, Aug 05, n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/737529948?accountid=26720.

[117] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[118] These are the concerns of Anthony Grafton, a historian at Princeton and president of the American Historical Association Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

[119] Geoffrey Nunberg Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[120] Siva Vaidhyanathan's concerns found here: Parry, Marc. "Google Starts Grant Program for Studies of Its Digitized Books." *Chronicle of Higher Education*, March 31, 2010. http://chronicle.com/article/Google-Starts-Grant-Program/64891/

[121] These are the concerns of Katie Trumpener, a professor of comparative literature and English at Yale University.

More from her: "If the whole field did that, that would be a disaster," she says, one that could yield a slew of insignificant numbers with "jumped-up claims about what they mean."

Novels are deeply specific, she argues, and the field has traditionally valued brilliant interpreters who create complex arguments about how that specificity works. When you treat novels as statistics, she says, the results can be misleading, because the reality of what you might include as a novel or what constitutes a genre is more slippery than a crude numerical picture can portray.

And then there's the question of whether transferring the lab model to a discipline like literary studies really works. Ms. Trumpener is dubious. Twenty postdocs carrying out one person's vision? She fears an "academia on autopilot," generating lots of research "without necessarily sharp critical intelligences guiding every phase of it." Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[122] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[123] Hand, Eric. "Culturonomics: Word Play," Nature 474, June 17, 2011 (published online): 436-440. http://www.nature.com/news/2011/110617/full/474436a.html , doi:10.1038/474436a

[124] Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[125] Cohen, Patricia. 2010. "Google Database Puts Language in a Petri Dish." *International Herald Tribune*, Dec 18, 12. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/818453126?accountid=26720

[126] "Google Book Privacy Still a Concern Post GBS," *Open Book Alliance* (blog), October 27, 2011. http://www.openbookalliance.org/2011/04/google-book-privacy-still-a-concern-post-gbs/

[127] "Hide from Google", *Wired How-to Wiki* (Wiki), Last modified: February 3, 2012 (10:30 p.m.) http://howto.wired.com/wiki/Hide_From_Google Also note the following quotes: "When you look back at your Google search history, it's like looking back through a log of your thoughts. Google has a better record of what you were thinking than you do…..What you were thinking (Google Search). What you were reading (Google Books). What you were talking about (Gmail, Google+). What you were watching (YouTube). And where you were you at the time (Google Maps)…..I've spent the last week researching this story on Google News, emailing sources through Gmail, and running Google searches for "Google privacy" and "Google lawsuit." Google knows exactly what I'm doing right now." Rowell, Rainbow, "Google is tracking your personal info", *Omaha World-Herald*, February 19, 2012. http://www.omaha.com/article/20120219/LIVING/702199995 Oh well? Another frightening quote: "More frightening to me still is Schirrmacher's postulated intelligent artificial agents who can, as in the Google Books example, search and access this knowledge base so quickly, and then integrate it to be used in real-time applications to manipulate the target individual to think or feel or behave in ways that suit the agent's (or its owner's) agenda of purposes. (Of course this is already being done in a crude way through advertising, both commercial and political; we have just shown for example that television snack food ads increase automatic consumption behavior in the viewer by nearly 50%, in children and adults alike." "The Age of the Informavore: A Talk with Frank Schirrmacher," *Edge*. October 27, 2009. http://www.edge.org/3rd_culture/schirrmacher09/schirrmacher09_index.html

[128] "Consider an interesting study of the titles of 19th-century books by the historians Dan Cohen and Fred Gibbs, of George Mason University, who also worked with the Google Books corpus. What does it signify that the words "hope" and "happiness" became less frequent in book titles in the second half of that century? To Cohen and Gibbs, it suggests that there was an undercurrent of depression during that period. But a reader of Schopenhauer might conclude that all those earlier mentions of happiness were the unmistakable signs of misery and abjection. To prove the case one way or the other, one might be driven

to, well, read some of the books." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[129] Haven, Cynthia, "Non-consumptive Research? Text Mining? Welcome to the Hotspot of Humanities Research at Stanford*" Stanford University News*, December 1, 2010. http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html

[130] Ideas for these formulations from this article: Ophir, Shai. "A New Type of Historical Knowledge." *Information Society* 26, no. 2 (March 2010): 144-150. *Academic Search Premier*, EBSCO*host* (accessed March 9, 2012).

[131] Haven, Cynthia, "Non-consumptive Research? Text Mining? Welcome to the Hotspot of Humanities Research at Stanford*" Stanford University News*, December 1, 2010. http://news.stanford.edu/news/2010/december/jockers-digitize-texts-120110.html

[132] On the other hand: "Mr. Jockers is less modest. In the lab, as the day winds down and chatter turns to what might be the next hot trend in literary studies, he taps his laptop and jackhammers his knee up and down. 'We're it,' he says." Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[133] On the other hand… : "One salutary effect of looking at word trajectories is that they dispel some of the unreflective philological assumptions that color the way humanists and social scientists tend to think about words." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/ ; Walt Crawford: "If you're suspicious that a clumsy plagiarist has cut-and-pasted without paraphrasing, almost any medium-length sentence may suggest you should check further. It could be entirely innocent. But it seems surprisingly uncommon for the same 10-word string to show up more than once. Our everyday language is more varied and diverse than I think most of us expect." Earlier, in his article, he had quoted a commenter on a blog who said, "it is highly likely that any given sentence you speak has never been used before, unless the sentence is short and about a common subject. It just seems like the same sentences get reused a lot because our brains are amazingly efficient at distilling sentences down to their core meanings, which do get reused regularly." In Crawford, Walt. "The Uniqueness of Everyday Language." *Online* 34, no. 4 (July 2010): 58-60. *Academic Search Premier*, EBSCO*host* (accessed March 9, 2012).

[134] Here are those points made in their fuller context: "I have a friend, a gifted amateur musician and computer scientist, who was involved in electronic music in its early days. Inevitably, within a few years, the field was taken over by composers. That happened partly because new interfaces made the technology more accessible, but also because a command of the subject matter always trumps mere technical expertise. As my friend put it, "It's a lot easier to turn an artist into a geek than to turn a geek into an artist."

 "Some people worry that the effect of these quantitative studies will be to trivialize scholarship. In a news article that appeared in *The Chronicle* last spring about Moretti's research, Katie Trumpener, a professor of comparative literature at Yale, voiced her concerns about the quantitative turn in literary studies. It's all well and good when it's done by an original thinker like Moretti, she said, but what happens when it's taken up by his "dullard" descendants? "If the whole field did that, that would be a disaster," with everyone producing insignificant numbers and "jumped-up claims about what they mean."

"It's unlikely that "the whole field" of literary studies—or any other field—will take up these methods, though the data will probably figure in the literature the way observations about origins and etymology do now. But I think Trumpener is quite right to predict that second-rate scholars will use the Google Books corpus to churn out gigabytes of uninformative graphs and insignificant conclusions. But it isn't as if those scholars would be doing more valuable work if they were approaching literature from some other point of view.

"This should reassure humanists about the immutably nonscientific status of their fields. Theories of what makes science science come and go, but one constant is that it proceeds by the aggregation of increments great and small, so that even the dullards have something to contribute. As William Whewell, who coined the word "scientist," put it, "Nothing which was done was useless or unessential." Humanists produce reams of work that is precisely that: useless because it's merely adequate. And the humanities resist the standardizations of method that make possible the structured collaborations of science, with the inevitable loss of individual voice. Whatever precedents yesterday's article in *Science* may establish for the humanities, the 12-author paper won't be one of them." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[135] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[136] Schumacher, E. F. *A Guide for the Perplexed*. New York: Harper & Row, 1977, 127.

[137] Parry, Marc. "The Humanities Go Google." *Chronicle of Higher Education*, May 28, 2010. http://chronicle.com/article/The-Humanities-Go-Google/65713/

[138] Musto, Ronald G. "Google Books Mutilates the Printed Past". *Chronicle of Higher Education.* 55, no. 39 (2009). He says at the end of his article: "If we acknowledge that Google Books is serving up to us only a mutilated, good-enough version of our already vicarious understanding of the past, what value does that hold for us? What dangers lie in wait for generations of students and scholars for whom the digital -- and Google's version of it -- will become the only reality? Must a whole new generation begin to reassemble the mutilations produced by Google Books to create authoritative and reliable digital texts? Must 2009 repeat the efforts of 1509 in reassembling, collating, editing, and republishing the scattered fragments of the manuscript past, which the age of print finally made uniform and authoritative? That would be absurd, precisely because it is so unnecessary. But should Google Books prevail, and the resources of the scholarly community be made irrelevant by Google's sheer scale and force, the future of our past will be in great doubt."

[139] "Early English Books Online (EEBO) published by ProQuest Information and Learning, Evans Early American Imprints from Newsbank-Readex, and Eighteenth Century Collections Online (ECCO) from Thompson-Gale". Martin, Shawn. "To Google or Not to Google, That is the Question: Supplementing Google Book Search to Make it More Useful for Scholarship." *Journal of Library Administration* 47, no 1-2 (2008): 141-150. More from Nunberg: "That leaves out a lot, compared with what you can do with other corpora. As of now, for example, you can't ask for a list of the words that follow the adjective "traditional" for each decade from 1900 to 2000 in order of descending frequency, or restrict a search for "*bronzino*" to paragraphs that contain "fish" and don't contain "painting." Some of those capabilities will

probably be available soon, though users won't be able to replicate many of the computationally heavy-duty exercises that the researchers report in the paper, and linguists won't really be happy until they can download the whole corpus and have their way with it.

And while the Harvard researchers have purged the research corpus of a large proportion of the metadata errors that have plagued Google Books, there are still a fair number of misdated works, and there's no way to restrict a query by genre or topic. You can ask the system to plot the trajectory "dear reader" in books published in Britain during the 19th century, but you can't limit the search to novels." Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[140] Nunberg, Geoffrey. "Counting on Google Books." *Chronicle of Higher Education*, December 16, 2010. http://chronicle.com/article/Counting-on-Google-Books/125735/

[141] See http://corpus.byu.edu/coha/compare-googleBooks.asp

[142] Martin, Shawn. "To Google or Not to Google, That is the Question: Supplementing Google Book Search to Make it More Useful for Scholarship." *Journal of Library Administration* 47, no 1-2 (2008): 141-150.

[143] As [the Center for Democracy and Technology] writes, Google:

"is moving right along with its project to digitize and offer books licensed through its Partner program. [And] [l]imiting customer previews, facilitate lending among friends, and sync "bookmarks" across devices all require fine-grained tracking of not just what books a person buys or browses, but also what pages she's read, what passages she's highlighted, and with whom she's shared which books. . . . Such a cache presents new opportunities for tracking and data mining, as well as a tempting honeypot for government or third-party litigants."

OBA has voiced the same types of concerns over in the past – especially in light of some of Google's highest ranking employees' views on privacy, like then-CEO Eric Schmidt's infamous statement that "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."

… The aggregation of data on what people are reading is a significant break with the centuries-old American tradition of vigorous protections of reader privacy. Back in February, we blogged about a paper written by two University of Washington professors on the privacy implications of Google's book project in which the authors write, "the fundamental goal of the American public library has for more than a century been to support the freedom of inquiry, and thereby the freedom of expression, necessary to the functioning of a free society." Certainly, reader privacy protections are needed in the Partner program as well as additional safeguards so that Google does not push the limits on tracking and data mining the sensitive information that CDT is concerned about…. "Google Book Privacy Still a Concern Post GBS," *Open Book Alliance* (blog), October 27, 2011. http://www.openbookalliance.org/2011/04/google-book-privacy-still-a-concern-post-gbs/

[144] "By stripping away many of the traditional safeguards on reader privacy— whether legal, ethical, or situational—shifting free-of-charge, publicly available reading from libraries to Google Books complicates the capacity of the context to support truly unfettered inquiry and knowledge diffusion. For all the reasons already noted—controversial interests, the ability of reading material to reveal other things

about the reader, or pure embarrassment—a lack of privacy with regard to the selection of reading materials can significantly chill individuals' desire and/or ability to explore as broadly as they might wish." Jones, Elisabeth A. and Janes, Joseph W. "Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read," *Policy & Internet*: 2, no. 4 (2010). http://www.psocommons.org/policyandinternet/vol2/iss4/art3/
DOI: 10.2202/1944-2866.1072

[145] "As part of the new policy Google will aggregate data it collects on users across its products, with the exception of Google Wallet and Google Books, and develop a mega-profile on each user. The data collection includes a user's Google searches, Gmail message content, contacts, YouTube favorites, and physical location."—Law, Ifrah, "EPIC Unlikely to Prevail in Challenge to FTC Stance on Google Privacy," JDSUPRA (blog), February 24, 2012. http://www.jdsupra.com/post/documentViewer.aspx?fid=29c4c5c1-4eec-4f14-8c9d-7b64a2dc3a87

[146] Desai, Santosh. "Column: Are Books on Google Good for Us?" *Financial Express*, (Feb 02, 2010) n/a. http://ezproxy.csp.edu/login?url=http://search.proquest.com/docview/872800719?accountid=26720.