May 14, 2006

Scan This Book!

By KEVIN KELLY

Correction Appended

In several dozen nondescript office buildings around the world, thousands of hourly workers bend over table-top scanners and haul dusty books into high-tech scanning booths. They are assembling the universal library page by page.

The dream is an old one: to have in one place all knowledge, past and present. All books, all documents, all conceptual works, in all languages. It is a familiar hope, in part because long ago we briefly built such a library. The great library at Alexandria, constructed around 300 B.C., was designed to hold all the scrolls circulating in the known world. At one time or another, the library held about half a million scrolls, estimated to have been between 30 and 70 percent of all books in existence then. But even before this great library was lost, the moment when all knowledge could be housed in a single building had passed. Since then, the constant expansion of information has overwhelmed our capacity to contain it. For 2,000 years, the universal library, together with other perennial longings like invisibility cloaks, antigravity shoes and paperless offices, has been a mythical dream that kept receding further into the infinite future.

Until now. When Google announced in December 2004 that it would digitally scan the books of five major research libraries to make their contents searchable, the promise of a universal library was resurrected. Indeed, the explosive rise of the Web, going from nothing to everything in one decade, has encouraged us to believe in the impossible again. Might the long-heralded great library of all knowledge really be within our grasp?

Brewster Kahle, an archivist overseeing another scanning project, says that the universal library is now within reach. "This is our chance to one-up the Greeks!" he shouts. "It is really possible with the technology of today, not tomorrow. We can provide all the works of humankind to all the people of the world. It will be an achievement remembered for all time, like putting a man on the moon." And unlike the libraries of old, which were restricted to the elite, this library would be truly democratic, offering every book to every person.

But the technology that will bring us a planetary source of all written material will also, in the same gesture, transform the nature of what we now call the book and the libraries that hold them. The universal library and its "books" will be unlike any library or books we have known. Pushing us rapidly toward that Eden of everything, and away from the paradigm of the physical paper tome, is the hot technology of the search engine.

1. Scanning the Library of Libraries

Scanning technology has been around for decades, but digitized books didn't make much sense until recently, when search engines like Google, Yahoo, Ask and MSN came along. When millions of books have been

scanned and their texts are made available in a single database, search technology will enable us to grab and read any book ever written. Ideally, in such a complete library we should also be able to read any article ever written in any newspaper, magazine or journal. And why stop there? The universal library should include a copy of every painting, photograph, film and piece of music produced by all artists, present and past. Still more, it should include all radio and television broadcasts. Commercials too. And how can we forget the Web? The grand library naturally needs a copy of the billions of dead Web pages no longer online and the tens of millions of blog posts now gone — the ephemeral literature of our time. In short, the entire works of humankind, from the beginning of recorded history, in all languages, available to all people, all the time.

This is a very big library. But because of digital technology, you'll be able to reach inside it from almost any device that sports a screen. From the days of Sumerian clay tablets till now, humans have "published" at least 32 million books, 750 million articles and essays, 25 million songs, 500 million images, 500,000 movies, 3 million videos, TV shows and short films and 100 billion public Web pages. All this material is currently contained in all the libraries and archives of the world. When fully digitized, the whole lot could be compressed (at current technological rates) onto 50 petabyte hard disks. Today you need a building about the size of a small-town library to house 50 petabytes. With tomorrow's technology, it will all fit onto your iPod. When that happens, the library of all libraries will ride in your purse or wallet — if it doesn't plug directly into your brain with thin white cords. Some people alive today are surely hoping that they die before such things happen, and others, mostly the young, want to know what's taking so long. (Could we get it up and running by next week? They have a history project due.)

Technology accelerates the migration of all we know into the universal form of digital bits. Nikon will soon quit making film cameras for consumers, and Minolta already has: better think digital photos from now on. Nearly 100 percent of all contemporary recorded music has already been digitized, much of it by fans. About one-tenth of the 500,000 or so movies listed on the Internet Movie Database are now digitized on DVD. But because of copyright issues and the physical fact of the need to turn pages, the digitization of books has proceeded at a relative crawl. At most, one book in 20 has moved from analog to digital. So far, the universal library is a library without many books.

But that is changing very fast. Corporations and libraries around the world are now scanning about a million books per year. Amazon has digitized several hundred thousand contemporary books. In the heart of Silicon Valley, Stanford University (one of the five libraries collaborating with Google) is scanning its eight-million-book collection using a state-of-the art robot from the Swiss company 4DigitalBooks. This machine, the size of a small S.U.V., automatically turns the pages of each book as it scans it, at the rate of 1,000 pages per hour. A human operator places a book in a flat carriage, and then pneumatic robot fingers flip the pages — delicately enough to handle rare volumes — under the scanning eyes of digital cameras.

Like many other functions in our global economy, however, the real work has been happening far away, while we sleep. We are outsourcing the scanning of the universal library. Superstar, an entrepreneurial company based in Beijing, has scanned every book from 900 university libraries in China. It has already digitized 1.3 million unique titles in Chinese, which it estimates is about half of all the books published in the Chinese language since 1949. It costs $30 to scan a book at Stanford but only $10 in China.

Raj Reddy, a professor at Carnegie Mellon University, decided to move a fair-size English-language library to

where the cheap subsidized scanners were. In 2004, he borrowed 30,000 volumes from the storage rooms of the Carnegie Mellon library and the Carnegie Library and packed them off to China in a single shipping container to be scanned by an assembly line of workers paid by the Chinese. His project, which he calls the Million Book Project, is churning out 100,000 pages per day at 20 scanning stations in India and China. Reddy hopes to reach a million digitized books in two years.

The idea is to seed the bookless developing world with easily available texts. Superstar sells copies of books it scans back to the same university libraries it scans from. A university can expand a typical 60,000-volume library into a 1.3 million-volume one overnight. At about 50 cents per digital book acquired, it's a cheap way for a library to increase its collection. Bill McCoy, the general manager of Adobe's e-publishing business, says: "Some of us have thousands of books at home, can walk to wonderful big-box bookstores and well-stocked libraries and can get Amazon.com to deliver next day. The most dramatic effect of digital libraries will be not on us, the well-booked, but on the billions of people worldwide who are underserved by ordinary paper books." It is these underbooked — students in Mali, scientists in Kazakhstan, elderly people in Peru — whose lives will be transformed when even the simplest unadorned version of the universal library is placed in their hands.

2. What Happens When Books Connect

The least important, but most discussed, aspects of digital reading have been these contentious questions: Will we give up the highly evolved technology of ink on paper and instead read on cumbersome machines? Or will we keep reading our paperbacks on the beach? For now, the answer is yes to both. Yes, publishers have lost millions of dollars on the long-prophesied e-book revolution that never occurred, while the number of physical books sold in the world each year continues to grow. At the same time, there are already more than a half a billion PDF documents on the Web that people happily read on computers without printing them out, and still more people now spend hours watching movies on microscopic cellphone screens. The arsenal of our current display technology — from handheld gizmos to large flat screens — is already good enough to move books to their next stage of evolution: a full digital scan.

Yet the common vision of the library's future (even the e-book future) assumes that books will remain isolated items, independent from one another, just as they are on shelves in your public library. There, each book is pretty much unaware of the ones next to it. When an author completes a work, it is fixed and finished. Its only movement comes when a reader picks it up to animate it with his or her imagination. In this vision, the main advantage of the coming digital library is portability — the nifty translation of a book's full text into bits, which permits it to be read on a screen anywhere. But this vision misses the chief revolution birthed by scanning books: in the universal library, no book will be an island.

Turning inked letters into electronic dots that can be read on a screen is simply the first essential step in creating this new library. The real magic will come in the second act, as each word in each book is cross-linked, clustered, cited, extracted, indexed, analyzed, annotated, remixed, reassembled and woven deeper into the culture than ever before. In the new world of books, every bit informs another; every page reads all the other pages.

In recent years, hundreds of thousands of enthusiastic amateurs have written and cross-referenced an entire

online encyclopedia called Wikipedia. Buoyed by this success, many nerds believe that a billion readers can reliably weave together the pages of old books, one hyperlink at a time. Those with a passion for a special subject, obscure author or favorite book will, over time, link up its important parts. Multiply that simple generous act by millions of readers, and the universal library can be integrated in full, by fans for fans.

In addition to a link, which explicitly connects one word or sentence or book to another, readers will also be able to add tags, a recent innovation on the Web but already a popular one. A tag is a public annotation, like a keyword or category name, that is hung on a file, page, picture or song, enabling anyone to search for that file. For instance, on the photo-sharing site Flickr, hundreds of viewers will "tag" a photo submitted by another user with their own simple classifications of what they think the picture is about: "goat," "Paris," "goofy," "beach party." Because tags are user-generated, when they move to the realm of books, they will be assigned faster, range wider and serve better than out-of-date schemes like the Dewey Decimal System, particularly in frontier or fringe areas like nanotechnology or body modification.

The link and the tag may be two of the most important inventions of the last 50 years. They get their initial wave of power when we first code them into bits of text, but their real transformative energies fire up as ordinary users click on them in the course of everyday Web surfing, unaware that each humdrum click "votes" on a link, elevating its rank of relevance. You may think you are just browsing, casually inspecting this paragraph or that page, but in fact you are anonymously marking up the Web with bread crumbs of attention. These bits of interest are gathered and analyzed by search engines in order to strengthen the relationship between the end points of every link and the connections suggested by each tag. This is a type of intelligence common on the Web, but previously foreign to the world of books.

Once a book has been integrated into the new expanded library by means of this linking, its text will no longer be separate from the text in other books. For instance, today a serious nonfiction book will usually have a bibliography and some kind of footnotes. When books are deeply linked, you'll be able to click on the title in any bibliography or any footnote and find the actual book referred to in the footnote. The books referenced in that book's bibliography will themselves be available, and so you can hop through the library in the same way we hop through Web links, traveling from footnote to footnote to footnote until you reach the bottom of things.

Next come the words. Just as a Web article on, say, aquariums, can have some of its words linked to definitions of fish terms, any and all words in a digitized book can be hyperlinked to other parts of other books. Books, including fiction, will become a web of names and a community of ideas.

Search engines are transforming our culture because they harness the power of relationships, which is all links really are. There are about 100 billion Web pages, and each page holds, on average, 10 links. That's a trillion electrified connections coursing through the Web. This tangle of relationships is precisely what gives the Web its immense force. The static world of book knowledge is about to be transformed by the same elevation of relationships, as each page in a book discovers other pages and other books. Once text is digital, books seep out of their bindings and weave themselves together. The collective intelligence of a library allows us to see things we can't see in a single, isolated book.

When books are digitized, reading becomes a community activity. Bookmarks can be shared with fellow

readers. Marginalia can be broadcast. Bibliographies swapped. You might get an alert that your friend Carl has annotated a favorite book of yours. A moment later, his links are yours. In a curious way, the universal library becomes one very, very, very large single text: the world's only book.

3. Books: The Liquid Version

At the same time, once digitized, books can be unraveled into single pages or be reduced further, into snippets of a page. These snippets will be remixed into reordered books and virtual bookshelves. Just as the music audience now juggles and reorders songs into new albums (or "playlists," as they are called in iTunes), the universal library will encourage the creation of virtual "bookshelves" — a collection of texts, some as short as a paragraph, others as long as entire books, that form a library shelf's worth of specialized information. And as with music playlists, once created, these "bookshelves" will be published and swapped in the public commons. Indeed, some authors will begin to write books to be read as snippets or to be remixed as pages. The ability to purchase, read and manipulate individual pages or sections is surely what will drive reference books (cookbooks, how-to manuals, travel guides) in the future. You might concoct your own "cookbook shelf" of Cajun recipes compiled from many different sources; it would include Web pages, magazine clippings and entire Cajun cookbooks. Amazon currently offers you a chance to publish your own bookshelves (Amazon calls them "listmanias") as annotated lists of books you want to recommend on a particular esoteric subject. And readers are already using Google Book Search to round up minilibraries on a certain topic — all books about Sweden, for instance, or books on clocks. Once snippets, articles and pages of books become ubiquitous, shuffle-able and transferable, users will earn prestige and perhaps income for curating an excellent collection.

Libraries (as well as many individuals) aren't eager to relinquish ink-on-paper editions, because the printed book is by far the most durable and reliable backup technology we have. Printed books require no mediating device to read and thus are immune to technological obsolescence. Paper is also extremely stable, compared with, say, hard drives or even CD's. In this way, the stability and fixity of a bound book is a blessing. It sits there unchanging, true to its original creation. But it sits alone.

So what happens when all the books in the world become a single liquid fabric of interconnected words and ideas? Four things: First, works on the margins of popularity will find a small audience larger than the near-zero audience they usually have now. Far out in the "long tail" of the distribution curve — that extended place of low-to-no sales where most of the books in the world live — digital interlinking will lift the readership of almost any title, no matter how esoteric. Second, the universal library will deepen our grasp of history, as every original document in the course of civilization is scanned and cross-linked. Third, the universal library of all books will cultivate a new sense of authority. If you can truly incorporate all texts — past and present, multilingual — on a particular subject, then you can have a clearer sense of what we as a civilization, a species, do know and don't know. The white spaces of our collective ignorance are highlighted, while the golden peaks of our knowledge are drawn with completeness. This degree of authority is only rarely achieved in scholarship today, but it will become routine.

Finally, the full, complete universal library of all works becomes more than just a better Ask Jeeves. Search on the Web becomes a new infrastructure for entirely new functions and services. Right now, if you mash up Google Maps and Monster.com, you get maps of where jobs are located by salary. In the same way, it is easy to see that in the great library, everything that has ever been written about, for example, Trafalgar Square in

London could be present on that spot via a screen. In the same way, every object, event or location on earth would "know" everything that has ever been written about it in any book, in any language, at any time. From this deep structuring of knowledge comes a new culture of interaction and participation.

The main drawback of this vision is a big one. So far, the universal library lacks books. Despite the best efforts of bloggers and the creators of the Wikipedia, most of the world's expertise still resides in books. And a universal library without the contents of books is no universal library at all.

There are dozens of excellent reasons that books should quickly be made part of the emerging Web. But so far they have not been, at least not in great numbers. And there is only one reason: the hegemony of the copy.

4. The Triumph of the Copy

The desire of all creators is for their works to find their way into all minds. A text, a melody, a picture or a story succeeds best if it is connected to as many ideas and other works as possible. Ideally, over time a work becomes so entangled in a culture that it appears to be inseparable from it, in the way that the Bible, Shakespeare's plays, "Cinderella" and the Mona Lisa are inseparable from ours. This tendency for creative ideas to infiltrate other works is great news for culture. In fact, this commingling of creations is culture.

In preindustrial times, exact copies of a work were rare for a simple reason: it was much easier to make your own version of a creation than to duplicate someone else's exactly. The amount of energy and attention needed to copy a scroll exactly, word for word, or to replicate a painting stroke by stroke exceeded the cost of paraphrasing it in your own style. So most works were altered, and often improved, by the borrower before they were passed on. Fairy tales evolved mythic depth as many different authors worked on them and as they migrated from spoken tales to other media (theater, music, painting). This system worked well for audiences and performers, but the only way for most creators to earn a living from their works was through the support of patrons.

That ancient economics of creation was overturned at the dawn of the industrial age by the technologies of mass production. Suddenly, the cost of duplication was lower than the cost of appropriation. With the advent of the printing press, it was now cheaper to print thousands of exact copies of a manuscript than to alter one by hand. Copy makers could profit more than creators. This imbalance led to the technology of copyright, which established a new order. Copyright bestowed upon the creator of a work a temporary monopoly — for 14 years, in the United States — over any copies of the work. The idea was to encourage authors and artists to create yet more works that could be cheaply copied and thus fill the culture with public works.

Not coincidentally, public libraries first began to flourish with the advent of cheap copies. Before the industrial age, libraries were primarily the property of the wealthy elite. With mass production, every small town could afford to put duplicates of the greatest works of humanity on wooden shelves in the village square. Mass access to public-library books inspired scholarship, reviewing and education, activities exempted in part from the monopoly of copyright in the United States because they moved creative works toward the public commons sooner, weaving them into the fabric of common culture while still remaining under the author's copyright. These are now known as "fair uses."

This wonderful balance was undone by good intentions. The first was a new copyright law passed by Congress

in 1976. According to the new law, creators no longer had to register or renew copyright; the simple act of creating something bestowed it with instant and automatic rights. By default, each new work was born under private ownership rather than in the public commons. At first, this reversal seemed to serve the culture of creation well. All works that could be copied gained instant and deep ownership, and artists and authors were happy. But the 1976 law, and various revisions and extensions that followed it, made it extremely difficult to move a work into the public commons, where human creations naturally belong and were originally intended to reside. As more intellectual property became owned by corporations rather than by individuals, those corporations successfully lobbied Congress to keep extending the once-brief protection enabled by copyright in order to prevent works from returning to the public domain. With constant nudging, Congress moved the expiration date from 14 years to 28 to 42 and then to 56.

While corporations and legislators were moving the goal posts back, technology was accelerating forward. In Internet time, even 14 years is a long time for a monopoly; a monopoly that lasts a human lifetime is essentially an eternity. So when Congress voted in 1998 to extend copyright an additional 70 years beyond the life span of a creator — to a point where it could not possibly serve its original purpose as an incentive to keep that creator working — it was obvious to all that copyright now existed primarily to protect a threatened business model. And because Congress at the same time tacked a 20-year extension onto all existing copyrights, nothing — no published creative works of any type — will fall out of protection and return to the public domain until 2019. Almost everything created today will not return to the commons until the next century. Thus the stream of shared material that anyone can improve (think "A Thousand and One Nights" or "Amazing Grace" or "Beauty and the Beast") will largely dry up.

In the world of books, the indefinite extension of copyright has had a perverse effect. It has created a vast collection of works that have been abandoned by publishers, a continent of books left permanently in the dark. In most cases, the original publisher simply doesn't find it profitable to keep these books in print. In other cases, the publishing company doesn't know whether it even owns the work, since author contracts in the past were not as explicit as they are now. The size of this abandoned library is shocking: about 75 percent of all books in the world's libraries are orphaned. Only about 15 percent of all books are in the public domain. A luckier 10 percent are still in print. The rest, the bulk of our universal library, is dark.

5. The Moral Imperative to Scan

The 15 percent of the world's 32 million cataloged books that are in the public domain are freely available for anyone to borrow, imitate, publish or copy wholesale. Almost the entire current scanning effort by American libraries is aimed at this 15 percent. The Million Book Project mines this small sliver of the pie, as does Google. Because they are in the commons, no law hinders this 15 percent from being scanned and added to the universal library.

The approximately 10 percent of all books actively in print will also be scanned before long. Amazon carries at least four million books, which includes multiple editions of the same title. Amazon is slowly scanning all of them. Recently, several big American publishers have declared themselves eager to move their entire backlist of books into the digital sphere. Many of them are working with Google in a partnership program in which Google scans their books, offers sample pages (controlled by the publisher) to readers and points readers to where they can buy the actual book. No one doubts electronic books will make money eventually. Simple

commercial incentives guarantee that all in-print and backlisted books will before long be scanned into the great library. That's not the problem.

The major problem for large publishers is that they are not certain what they actually own. If you would like to amuse yourself, pick an out-of-print book from the library and try to determine who owns its copyright. It's not easy. There is no list of copyrighted works. The Library of Congress does not have a catalog. The publishers don't have an exhaustive list, not even of their own imprints (though they say they are working on it). The older, the more obscure the work, the less likely a publisher will be able to tell you (that is, if the publisher still exists) whether the copyright has reverted to the author, whether the author is alive or dead, whether the copyright has been sold to another company, whether the publisher still owns the copyright or whether it plans to resurrect or scan it. Plan on having a lot of spare time and patience if you inquire. I recently spent two years trying to track down the copyright to a book that led me to Random House. Does the company own it? Can I reproduce it? Three years later, the company is still working on its answer. The prospect of tracking down the copyright — with any certainty — of the roughly 25 million orphaned books is simply ludicrous.

Which leaves 75 percent of the known texts of humans in the dark. The legal limbo surrounding their status as copies prevents them from being digitized. No one argues that these are all masterpieces, but there is history and context enough in their pages to not let them disappear. And if they are not scanned, they in effect will disappear. But with copyright hyperextended beyond reason (the Supreme Court in 2003 declared the law dumb but not unconstitutional), none of this dark library will return to the public domain (and be cleared for scanning) until at least 2019. With no commercial incentive to entice uncertain publishers to pay for scanning these orphan works, they will vanish from view. According to Peter Brantley, director of technology for the California Digital Library, "We have a moral imperative to reach out to our library shelves, grab the material that is orphaned and set it on top of scanners."

No one was able to unravel the Gordian knot of copydom until 2004, when Google came up with a clever solution. In addition to scanning the 15 percent out-of-copyright public-domain books with their library partners and the 10 percent in-print books with their publishing partners, Google executives declared that they would also scan the 75 percent out-of-print books that no one else would touch. They would scan the entire book, without resolving its legal status, which would allow the full text to be indexed on Google's internal computers and searched by anyone. But the company would show to readers only a few selected sentence-long snippets from the book at a time. Google's lawyers argued that the snippets the company was proposing were something like a quote or an excerpt in a review and thus should qualify as a "fair use."

Google's plan was to scan the full text of every book in five major libraries: the more than 10 million titles held by Stanford, Harvard, Oxford, the University of Michigan and the New York Public Library. Every book would be indexed, but each would show up in search results in different ways. For out-of-copyright books, Google would show the whole book, page by page. For the in-print books, Google would work with publishers and let them decide what parts of their books would be shown and under what conditions. For the dark orphans, Google would show only limited snippets. And any copyright holder (author or corporation) who could establish ownership of a supposed orphan could ask Google to remove the snippets for any reason.

At first glance, it seemed genius. By scanning all books (something only Google had the cash to do), the

company would advance its mission to organize all knowledge. It would let books be searchable, and it could potentially sell ads on those searches, although it does not do that currently. In the same stroke, Google would rescue the lost and forgotten 75 percent of the library. For many authors, this all-out campaign was a salvation. Google became a discovery tool, if not a marketing program. While a few best-selling authors fear piracy, every author fears obscurity. Enabling their works to be found in the same universal search box as everything else in the world was good news for authors and good news for an industry that needed some. For authors with books in the publisher program and for authors of books abandoned by a publisher, Google unleashed a chance that more people would at least read, and perhaps buy, the creation they had sweated for years to complete.

6. The Case Against Google

Some authors and many publishers found more evil than genius in Google's plan. Two points outraged them: the virtual copy of the book that sat on Google's indexing server and Google's assumption that it could scan first and ask questions later. On both counts the authors and publishers accused Google of blatant copyright infringement. When negotiations failed last fall, the Authors Guild and five big publishing companies sued Google. Their argument was simple: Why shouldn't Google share its ad revenue (if any) with the copyright owners? And why shouldn't Google have to ask permission from the legal copyright holder before scanning the work in any case? (I have divided loyalties in the case. The current publisher of my books is suing Google to protect my earnings as an author. At the same time, I earn income from Google Adsense ads placed on my blog.)

One mark of the complexity of this issue is that the publishers suing were, and still are, committed partners in the Google Book Search Partner Program. They still want Google to index and search their in-print books, even when they are scanning the books themselves, because, they say, search is a discovery tool for readers. The ability to search the scans of all books is good for profits.

The argument about sharing revenue is not about the three or four million books that publishers care about and keep in print, because Google is sharing revenues for those books with publishers. (Google says publishers receive the "majority share" of the income from the small ads placed on partner-program pages.) The argument is about the 75 percent of books that have been abandoned by publishers as uneconomical. One curious fact, of course, is that publishers only care about these orphans now because Google has shifted the economic equation; because of Book Search, these dark books may now have some sparks in them, and the publishers don't want this potential revenue stream to slip away from them. They are now busy digging deep into their records to see what part of the darkness they can declare as their own.

The second complaint against Google is more complex. Google argues that it is nearly impossible to track down copyright holders of orphan works, and so, it says, it must scan those books first and only afterward honor any legitimate requests to remove the scan. In this way, Google follows the protocol of the Internet. Google scans all Web pages; if it's on the Web, it's scanned. Web pages, by default, are born copyrighted. Google, therefore, regularly copies billions of copyrighted pages into its index for the public to search. But if you don't want Google to search your Web site, you can stick some code on your home page with a no-searching sign, and Google and every other search engine will stay out. A Web master thus can opt out of search. (Few do.) Google applies the same principle of opting-out to Book Search. It is up to you as an author

to notify Google if you don't want the company to scan or search your copyrighted material. This might be a reasonable approach for Google to demand from an author or publisher if Google were the only search company around. But search technology is becoming a commodity, and if it turns out there is any money in it, it is not impossible to imagine a hundred mavericks scanning out-of-print books. Should you as a creator be obliged to find and notify each and every geek who scanned your work, if for some reason you did not want it indexed? What if you miss one?

There is a technical solution to this problem: for the search companies to compile and maintain a common list of no-scan copyright holders. A publisher or author who doesn't want a work scanned notifies the keepers of the common list once, and anyone conducting scanning would have to remove material that was listed. Since Google, like all the other big search companies — Microsoft, Amazon and Yahoo — is foremost a technical-solution company, it favors this approach. But the battle never got that far.

7. When Business Models Collide

In thinking about the arguments around search, I realized that there are many ways to conceive of this conflict. At first, I thought that this was a misunderstanding between people of the book, who favor solutions by laws, and people of the screen, who favor technology as a solution to all problems. Last November, the New York Public Library (one of the "Google Five") sponsored a debate between representatives of authors and publishers and supporters of Google. I was tickled to see that up on the stage, the defenders of the book were from the East Coast and the defenders of the screen were from the West Coast. But while it's true that there's a strand of cultural conflict here, I eventually settled on a different framework, one that I found more useful. This is a clash of business models.

Authors and publishers (including publishers of music and film) have relied for years on cheap mass-produced copies protected from counterfeits and pirates by a strong law based on the dominance of copies and on a public educated to respect the sanctity of a copy. This model has, in the last century or so, produced the greatest flowering of human achievement the world has ever seen, a magnificent golden age of creative works. Protected physical copies have enabled millions of people to earn a living directly from the sale of their art to the audience, without the weird dynamics of patronage. Not only did authors and artists benefit from this model, but the audience did, too. For the first time, billions of ordinary people were able to come in regular contact with a great work. In Mozart's day, few people ever heard one of his symphonies more than once. With the advent of cheap audio recordings, a barber in Java could listen to them all day long.

But a new regime of digital technology has now disrupted all business models based on mass-produced copies, including individual livelihoods of artists. The contours of the electronic economy are still emerging, but while they do, the wealth derived from the old business model is being spent to try to protect that old model, through legislation and enforcement. Laws based on the mass-produced copy artifact are being taken to the extreme, while desperate measures to outlaw new technologies in the marketplace "for our protection" are introduced in misguided righteousness. (This is to be expected. The fact is, entire industries and the fortunes of those working in them are threatened with demise. Newspapers and magazines, Hollywood, record labels, broadcasters and many hard-working and wonderful creative people in those fields have to change the model of how they earn money. Not all will make it.)

The new model, of course, is based on the intangible assets of digital bits, where copies are no longer cheap but free. They freely flow everywhere. As computers retrieve images from the Web or display texts from a server, they make temporary internal copies of those works. In fact, every action you take on the Net or invoke on your computer requires a copy of something to be made. This peculiar superconductivity of copies spills out of the guts of computers into the culture of computers. Many methods have been employed to try to stop the indiscriminate spread of copies, including copy-protection schemes, hardware-crippling devices, education programs, even legislation, but all have proved ineffectual. The remedies are rejected by consumers and ignored by pirates.

As copies have been dethroned, the economic model built on them is collapsing. In a regime of superabundant free copies, copies lose value. They are no longer the basis of wealth. Now relationships, links, connection and sharing are. Value has shifted away from a copy toward the many ways to recall, annotate, personalize, edit, authenticate, display, mark, transfer and engage a work. Authors and artists can make (and have made) their livings selling aspects of their works other than inexpensive copies of them. They can sell performances, access to the creator, personalization, add-on information, the scarcity of attention (via ads), sponsorship, periodic subscriptions — in short, all the many values that cannot be copied. The cheap copy becomes the "discovery tool" that markets these other intangible valuables. But selling things-that-cannot-be-copied is far from ideal for many creative people. The new model is rife with problems (or opportunities). For one thing, the laws governing creating and rewarding creators still revolve around the now-fragile model of valuable copies.

8. Search Changes Everything

The search-engine companies, including Google, operate in the new regime. Search is a wholly new concept, not foreseen in version 1.0 of our intellectual-property law. In the words of a recent ruling by the United States District Court for Nevada, search has a "transformative purpose," adding new social value to what it searches. What search uncovers is not just keywords but also the inherent value of connection. While almost every artist recognizes that the value of a creation ultimately rests in the value he or she personally gets from creating it (and for a few artists that value is sufficient), it is also true that the value of any work is increased the more it is shared. The technology of search maximizes the value of a creative work by allowing a billion new connections into it, often a billion new connections that were previously inconceivable. Things can be found by search only if they radiate potential connections. These potential relationships can be as simple as a title or as deep as hyperlinked footnotes that lead to active pages, which are also footnoted. It may be as straightforward as a song published intact or as complex as access to the individual instrument tracks — or even individual notes.

Search opens up creations. It promotes the civic nature of publishing. Having searchable works is good for culture. It is so good, in fact, that we can now state a new covenant: Copyrights must be counterbalanced by copyduties. In exchange for public protection of a work's copies (what we call copyright), a creator has an obligation to allow that work to be searched. No search, no copyright. As a song, movie, novel or poem is searched, the potential connections it radiates seep into society in a much deeper way than the simple publication of a duplicated copy ever could.

We see this effect most clearly in science. Science is on a long-term campaign to bring all knowledge in the world into one vast, interconnected, footnoted, peer-reviewed web of facts. Independent facts, even those that

make sense in their own world, are of little value to science. (The pseudo- and parasciences are nothing less, in fact, than small pools of knowledge that are not connected to the large network of science.) In this way, every new observation or bit of data brought into the web of science enhances the value of all other data points. In science, there is a natural duty to make what is known searchable. No one argues that scientists should be paid when someone finds or duplicates their results. Instead, we have devised other ways to compensate them for their vital work. They are rewarded for the degree that their work is cited, shared, linked and connected in their publications, which they do not own. They are financed with extremely short-term (20-year) patent monopolies for their ideas, short enough to truly inspire them to invent more, sooner. To a large degree, they make their living by giving away copies of their intellectual property in one fashion or another.

The legal clash between the book copy and the searchable Web promises to be a long one. Jane Friedman, the C.E.O. of HarperCollins, which is supporting the suit against Google (while remaining a publishing partner), declared, "I don't expect this suit to be resolved in my lifetime." She's right. The courts may haggle forever as this complex issue works its way to the top. In the end, it won't matter; technology will resolve this discontinuity first. The Chinese scanning factories, which operate under their own, looser intellectual-property assumptions, will keep churning out digital books. And as scanning technology becomes faster, better and cheaper, fans may do what they did to music and simply digitize their own libraries.

What is the technology telling us? That copies don't count any more. Copies of isolated books, bound between inert covers, soon won't mean much. Copies of their texts, however, will gain in meaning as they multiply by the millions and are flung around the world, indexed and copied again. What counts are the ways in which these common copies of a creative work can be linked, manipulated, annotated, tagged, highlighted, bookmarked, translated, enlivened by other media and sewn together into the universal library. Soon a book outside the library will be like a Web page outside the Web, gasping for air. Indeed, the only way for books to retain their waning authority in our culture is to wire their texts into the universal library.

But the reign of livelihoods based on the copy is not over. In the next few years, lobbyists for book publishers, movie studios and record companies will exert every effort to mandate the extinction of the "indiscriminate flow of copies," even if it means outlawing better hardware. Too many creative people depend on the business model revolving around copies for it to pass quietly. For their benefit, copyright law will not change suddenly.

But it will adapt eventually. The reign of the copy is no match for the bias of technology. All new works will be born digital, and they will flow into the universal library as you might add more words to a long story. The great continent of orphan works, the 25 million older books born analog and caught between the law and users, will be scanned. Whether this vast mountain of dark books is scanned by Google, the Library of Congress, the Chinese or by readers themselves, it will be scanned well before its legal status is resolved simply because technology makes it so easy to do and so valuable when done. In the clash between the conventions of the book and the protocols of the screen, the screen will prevail. On this screen, now visible to one billion people on earth, the technology of search will transform isolated books into the universal library of all human knowledge.

Kevin Kelly is the "senior maverick" at Wired magazine and author of "Out of Control: The New Biology of Machines, Social Systems and the Economic World" and other books. He last wrote for the magazine about digital music.

Correction: May 14, 2006

An article on Page 42 of The Times Magazine today about the future of book publishing misstates the number and type of libraries in China from which a Chinese company, Superstar, has made digital copies of books. It is 200 libraries of all kinds, not 900 university libraries.

In fact, I've scanned a 600+ page history book in under 2 hours! In this instructable I will show you the methods I used to both capture every page of a book with a digital camera, and later process the images into a readable PDF file using cheap/free software. DISCLAIMER: This information is for educational purposes only. I am in no way responsible for anything illegal you do with this knowledge. In this article, we're going to review the best book scanners for the money. But, first things first. Scanning, at its core, takes away the need to fill paper forms. Even more impressive, it makes sharing files super-easy. Still, to get a model that suits your needs, you have to do some research – and that's where we come in. Put differently we've scoured the market in search of top-rated books scanners and what makes them special. And that way, you don't have to second-guess your choice. Click through to see what we mean. Scan this book by John Mendenhall, 1992, Art Direction Book Co. edition, in English. Donate this book to the Internet Archive library. ×—Close. Hooray! You've discovered a title that's missing from our library. Can you help donate a copy? If you own this book, you can mail it to our address below. You can also purchase this book from a vendor and ship it to our address: Internet Archive Open Library Book Donations 300 Funston Avenue San Francisco, CA 94118. Better World Books. Amazon. More. Bookshop.org. Scan This Book Three: Thousands More Copyright-Free Objects for Scanning by John Mendenhall Paperback $24.95. Only 1 left in stock - order soon. Ships from and sold by fleamarketbooks. Easy to lay flat on a scanner or photocopier, well organized with a handy index in the back. Multiple images of similar items that are clear and useful. I've been looking for a clip art book for a project I've been working on and this one has almost all the images I need. I can't say it'll be perfect for you but its a book well worth having. OCA book scan program is an | Find, read and cite all the research you need on ResearchGate. Scan This Book! January 2007. Library Journal 132(13). Authors: Andrew Richard Albanese. Request full-text PDF. To read the full-text of this research, you can request a copy directly from the author. Request full-text. Download citation.