

# Public Domain Blueprint

## Part 2b

Welcome to Part two of Week two of Public Domain Blueprint. This is the continuation of the previous webinar and in this one we are going to get into the down and dirty of finding books online.

I'm primarily going to focus on five websites and these are the ones I usually spend most of my time. However, with the bonus I'm including for this session, which is Public Domain Code Book, there's over 200 additional websites you can access to find books and a whole lot more. That is why I decided to include that as a bonus.

But I will tell you that the majority of my personal research time is spent focusing on the few sites I'm sharing with you. And these are the ones I suggest you focus on if you want to find books online or even find books that perhaps aren't online. In other words, books that are not currently available in digital form. They may not be online, but there are a couple of online resources we can use to discover those as well.

# Books.google.com

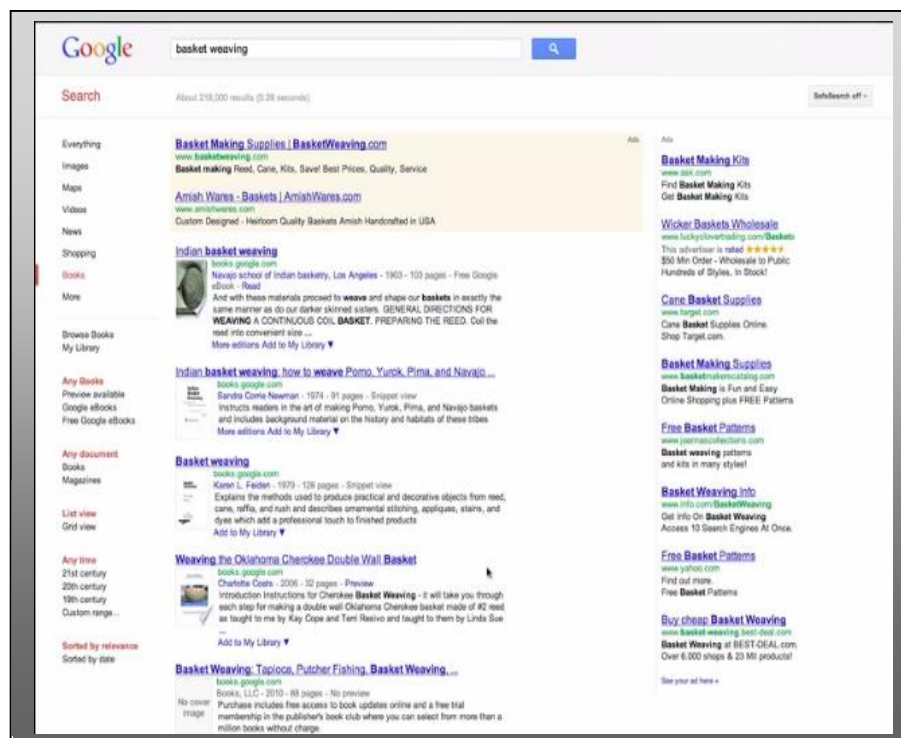
The first resource we will look at is pretty obvious, and it is Google Books. Google Books really represents the largest repository of digital books in the public domain right now. What I'm going to do, as we go through these five different resources, I'm going to give you a grand tour overview of each one of them and show you how to use them.

If you've never been there before there are a couple things you need to be aware of. Simply go to [books.google.com](http://books.google.com) and essentially what we are going to do is just pick a topic to do a search in so you can get a feel for how the process works.

Let's pick a topic I really enjoy and grew up doing, basket weaving. My mom's family is Cherokee so I grew up being taught Cherokee basket weaving by my mom and that was kind of fun so let's see what we can discover. We click

'search books' and you'll notice a large variety of books, many of them are not in the public domain of course. Here is one that says 2010, 1906, 2007...the

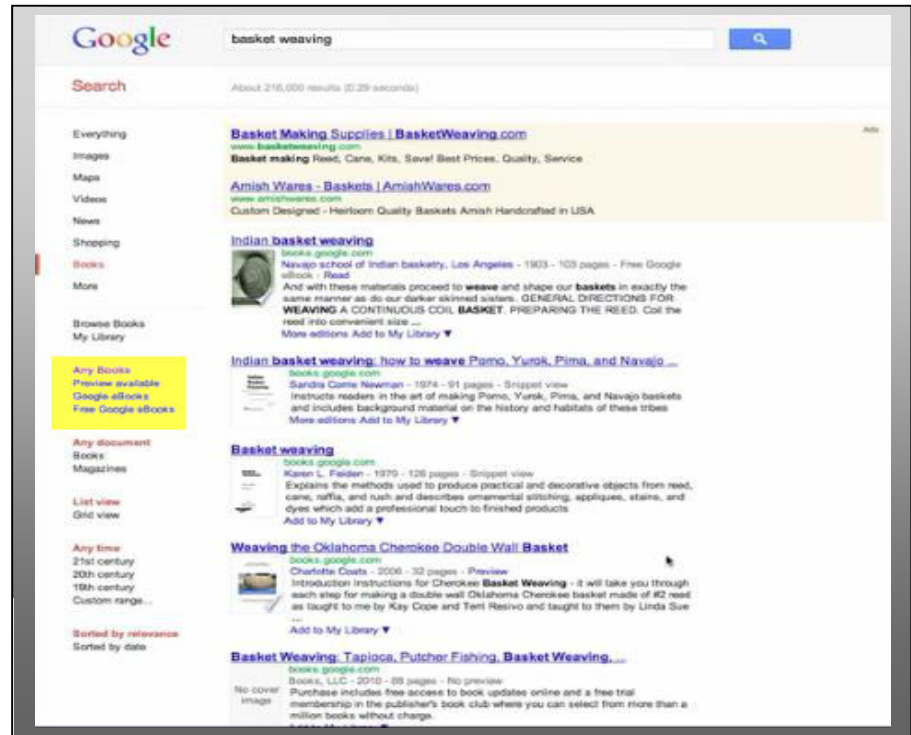
*Prefer to watch?  
Look over Tony's shoulder  
as he explores Google  
Books in video 2a,  
beginning at 1:13*



public domain books obviously mixed in and interwoven with books that are not in the public domain.

Google actually makes it extremely easy to separate those and the way you do that is come over here to where it says 'any books' and at the bottom where it says 'free Google e-books' you click that link. Instantly it removed all of the more current books and limits the search results to public domain books. You can see that just by the publish dates, which they include with all the different books.

We have a few results, like 35,000 which isn't too bad. So if we take the first

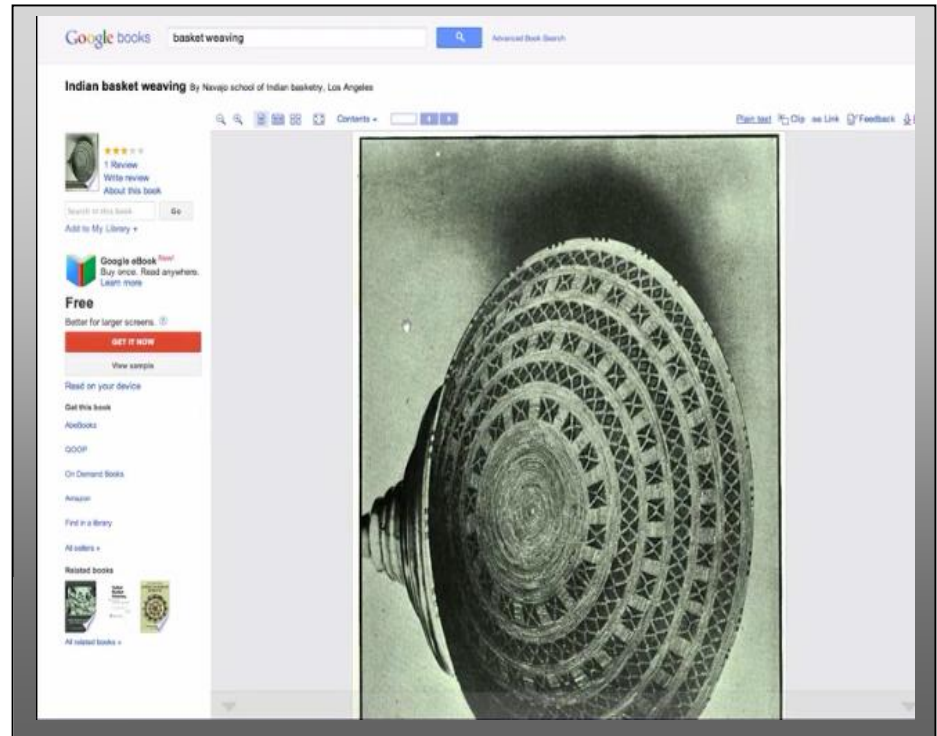


book here, Indian Basket Weaving and click on that, you can see that Google makes this funky make shift cover and then it is all straight forward scans. Very black and white and all that. You can search through this book because they run it through OCR which is good for us except if we want to download the entire book, it becomes a little more challenging. We will get into this process more later, but we have a few options. You can see here on the far right where it says we can download the PDF, if we click on it we can download the PDF, the entire book. Unfortunately, the embedded text is not included in that PDF, it is something Google thinks they are all clever about doing. They do provide a link for plain text and if we click on that essentially it takes the OCR that Google

has done and reveals it except it only does so in chunks. So you can see about five page chunks. The highlighted word is there because it is one of the keywords I used.

The OCR on the text isn't bad, they even include the images. The basket images are extremely small; they would never be usable if you wanted to turn this into a print book.

But the OCR itself is really well done and so the premise is that you could just highlight the text and copy/paste it into a Word document, five pages at a time. If you have a several hundred page book, that could be



intimidating. I don't know if Google allows you to change this, let's find out. I don't even know how many pages there are in this book, but let's try. Let's see if it will allow us to do this, they are kind of sneaky, they aren't letting us get away with anything. So, you're kind of limited as you can see. It would take a little bit of doing, probably somebody applying focused effort, if you have a couple hundred page book you could go through the five pages at a time and copy/paste the text out under half an hour which in the grand scheme of things really isn't that long. Now, there is another way and we will get into that way in a little bit.

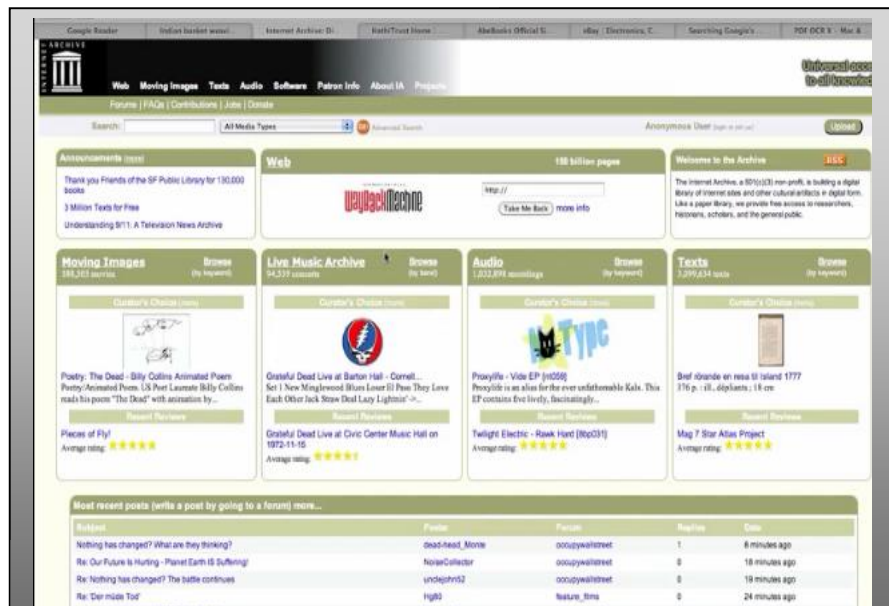
So, right now, we're going to move on to the next source. Again, it is going to be pretty straight forward and we'll see what we find.

# Archive.org

Explore archive.org with  
Tony on Video 2a  
beginning at 7:13

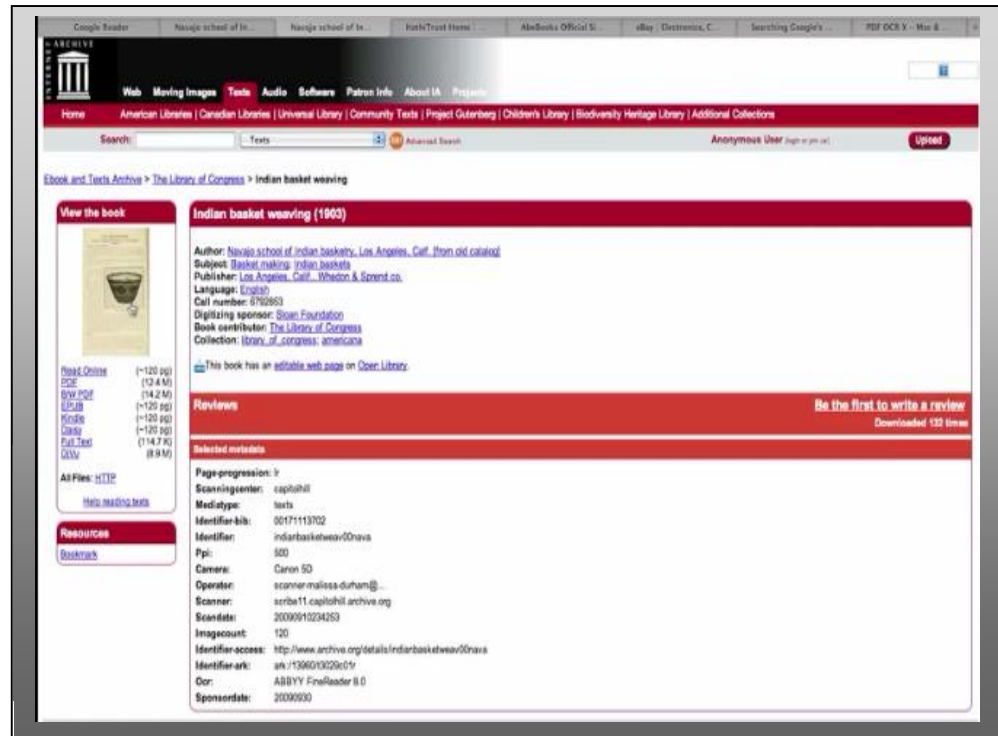
The next resource we will look at is Internet Archive. I like Internet Archive a lot because it takes Google Books to the next level and a lot of the books on Internet Archives are from Google Books. But, there are a number of volunteers who also take these books from Google and convert them into a variety of forms that are much more usable.

That is why I like





Internet Archive. If we, for the sake of discussion, go back here to our book, Indian Basket Weaving and see if we can find that on Internet Archive, again, just like before on Google, you type in your keyword phrase, which we are a little bit more focused now because we are looking for a specific book but I could type in 'basket weaving' as I did on Google and it would bring up the results. Here it says 'all media types' and we don't want that, we want to focus on text. So we just choose text, and click 'go.'



You can see here where there's actually a couple of choices, if we read it, it says Indian basket weaving, the first book digitized by Google but here's another one that it doesn't say that it was digitized by Google so let's take a look at this one. You can see here that you can read it online, you can download the PDF, you can download the full text and this one is actually different from the one we saw on Google because you can see here by the little preview pictures it not been converted to just strictly black and white, you can still see the higher dynamic range of images because they include the original paper color and all that fun stuff, which is a good thing. Just like I mentioned before in extracting text, Internet Archive have some very specific strategies that work really well in extracting both text and images. We will come back to that because I don't

want to keep jumping back and forth, I wanted to be a smooth process of actually getting the text, the best ways to do that and how to work with it and all of that. Will come back to that but I want to leave this particular page up but I again, it is very straightforward in locating this content and you'll notice I again they have the published year and everything, it's very straightforward which is a good thing.

# Hathitrust.org

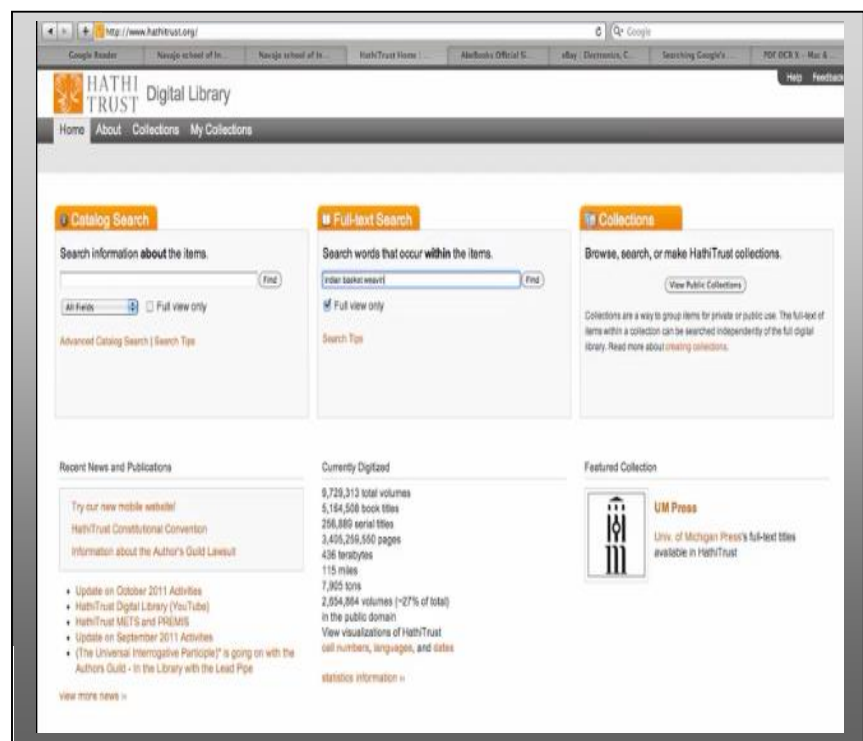
*Explore Hathi Trust by following along with Tony on video 2a, beginning at 10:08*

The next resource we are going to look at, and by the way, just these couple of resources, these to write here represents about 4 million books if not more. The third one we will look at is Hathi Trust, so let's flip over to that page.

Again and there is a huge digital library, in this case what we're interested

in is full text search. So it actually searches within the items, here you can search about the items so that is more of a title thing; this is more of a content thing. You always want to make sure that you choose full view only because they have books on here that are not full view, which means there's not complete copies that

you can download. Here it says there are 9.7 million total books in but that doesn't necessarily mean that all of those are in the public domain. Actually, it

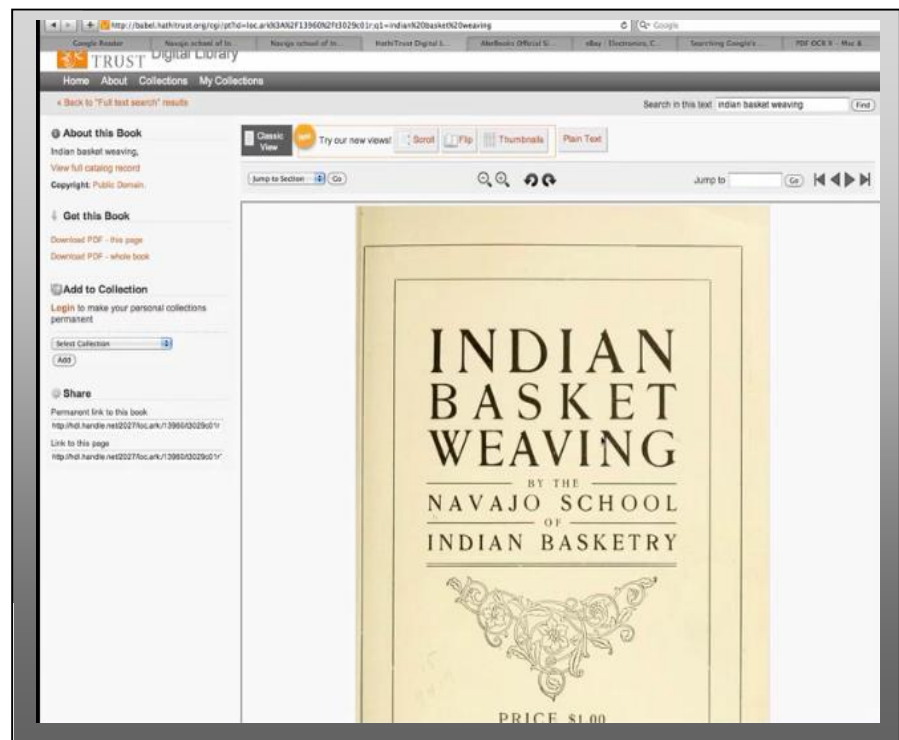


says right here, 2.6 million give or take 27% of the total are in the public domain. So there are roughly 2 million books are in the public domain, that's not bad, that kind of fun.

So let's type in, Indian basket weaving and what we want is the full view so we will click on that. So you can see here is the book and we can choose to zoom in if we want will. So you can see that the detail looks a lot better on the images of the basket than what we saw in Google by a lot.

We can also switch to plain text which in this particular side is page for page so we are talking worse than Google; at least Google shows five pages at a time. This site does not but you still can copy it.

You also have



the ability to download the entire PDF. There are some PDFs that you have to register on this site to download, which is free to do and it actually creates PDF on the fly. It builds them automatically, dynamically, which is why it is doing what it's doing right now. As you can see we are building that PDF but, again, like what we discovered with Google, the PDF we end up with is not editable, still making it extremely challenging to get that text. As soon as this is done building we will open it up just so you can see what I'm talking about. It actually builds it by page images, which isn't necessarily a problem because I have a fun solution

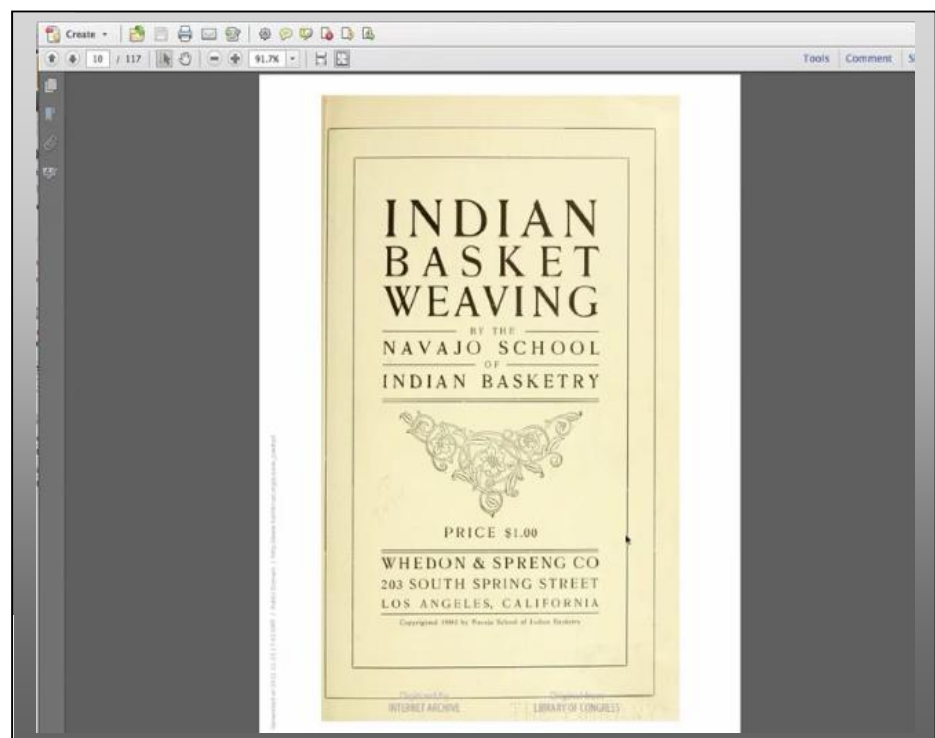


to show you here in a little bit. It's actually a new solution from anything I've talked before because it is a newer software.

We are almost done building here and these three websites, Google Books, Archive.org and HathiTrust.org, those really represent the largest volume of public domains books on the Internet. There are a couple others that are a pretty big repository but these are the three that I usually go to first just because of the sheer volume of content that they have, it makes sense. Some of the other websites you'll find in the Public Domain Code Book are actually more niche specific so if you have a specific interest in a certain topic, there are some websites out there that I found and identified that can make finding those resources much more targeted. For instant, one of the resources included in the book is exclusively devoted to fly fishing. Another one is devoted to home economics.

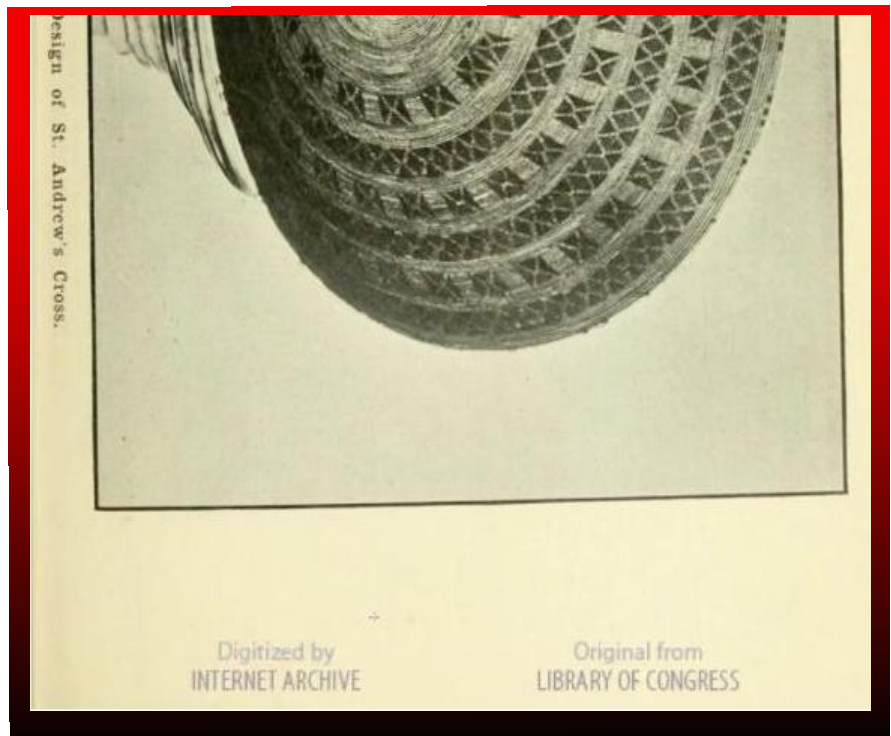
Let's download our PDF here, it's a pretty big PDF it is 77.5 MB. I haven't seen this PDF before by the way, so I don't know what it will look like. But I've downloaded PDFs from this site before which is why I've kind of have an idea what to expect.

Notice that is says 'copyright public domain' which is good. All right, let's get that out of the way. Here is our image, our PDF and you can see that it is the actual pages. If I click on this you



can see that I cannot highlight the text. Let's try zooming in just to see what kind of quality we have here. It starts to get fuzzy blurry at 400, is not quite as bad at 200. The reason I'm checking this is because, if I wanted to make a print book out of this, I need the best quality images of the baskets I can get. That's why I'm looking.

We can see here that we have text but it won't necessarily help us out so much. Check this out, notice where it says that it was digitized by Internet archive, a lot of these sites are swapping back and forth so even though Hathi Trust has 2.2 million books in the public domain it doesn't mean those 2 million books are different from Google Books and Internet Archive. There is a lot of overlap within these three sites.

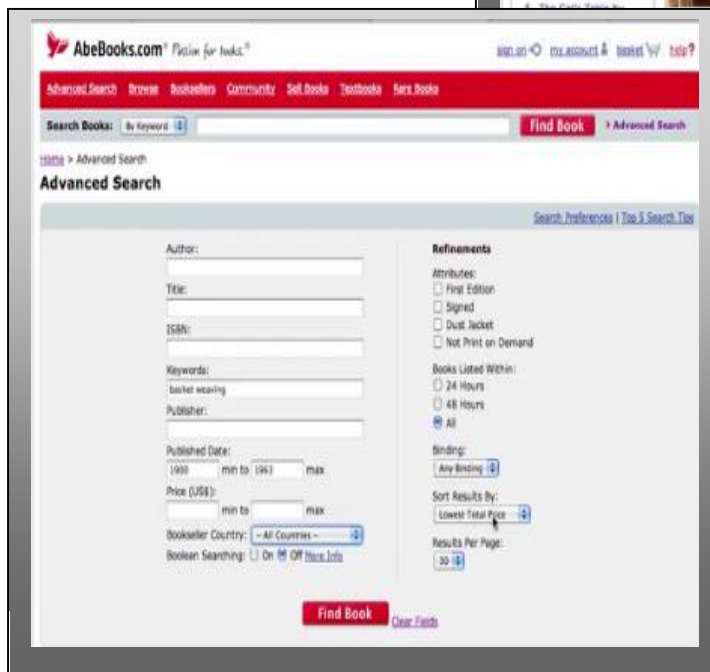


Alright, that's Hathi Trust. As you can see, it is very straightforward and pretty much works the way I have shown you, all the books work pretty much the same way.

# Abebooks.com

*Discover Abe Books on Video 2a,  
beginning at 17:37*

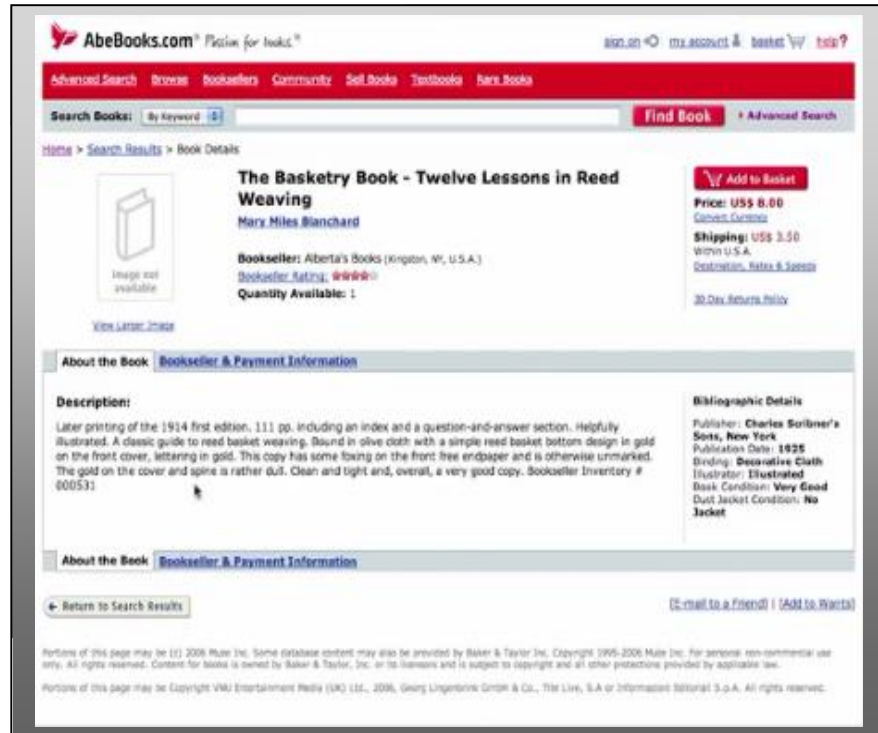
The next one we are going to look at isn't a digital repository but it is a good place to find those books that are in the public domain that may not be available online. One of the largest online resources for that is Abe Books, and actually apart from the next one I'm going to show you, which is eBay, Abe Books is pretty much the only place I go to buy used books. But there are some specific things you want to keep in mind as far as searching and finding what you want. What I usually do



is click on more search options which allows us to get much more specific.

If we want to search on a topic instead of a specific title, let's say I want to search on books on basket weaving in general. We'll type that in the keywords section and then we

can limit the published years. Say we want 1900 to, well the copyright laws changed in 1964, so let's say 1963. We can go with 'booksellers in all countries,' we don't need to worry about all this other stuff for right now but we will search the results by



lowest total price, we might as well get it as cheap as we can, right? So click 'find book', and now you'll notice it is actually bringing up magazines too, which is cool. But as we scroll down through here you'll notice we have a number of choices, essentially what we would want to do is if we identify a book we might be interested in buying, like here is that basketry book 12 Lesson in Reed Weaving which sounds interesting, like it might be a book. They don't have a picture of the cover but if we read the description it says that it is a later printing of the 1914 addition, question and answer sections, highly illustrated, classic guide. So it sounds like it might be perfect and between the price and shipping you are talking about \$11.50. Granted, we would have to scan it but its it's likely this book may not be in online at all, I haven't checked I have no idea, I just randomly chose this one. I'm sure someone who is watching this video will check that.

This is the 1925 edition so we would want to check and verify the copyright status even though the 1914 at dish and would definitely be in the public domain because it's before 1923. The 1925 addition could be renewed, it

is likely that it wasn't but it could be. You would just use the copyright research methods that I shared in the other chapter for copyright research and determine whether the book was in the public domain and I recommend you do that before you buy it. If you discover it is in the public domain, then buy it. Then you can either scan it yourself or you can outsource the scanning. There are a number of organizations that will scan books and will convert them into editable text. That is a definite option for you and it makes it pretty straightforward.

One of the things that I love about Abe Books is you have the ability to search 140 million books so it is likely that you may actually find books that you are looking for that are not available online but may contain great content. That is something good to know.

## eBay.com

*Discover the research power of eBay  
with Tony on Video 2a, beginning at  
22:07*

As I mentioned, the last resource I'm going to share here is eBay.com. Of course you all are familiar with eBay and we've been talking about eBay just with what I shared in the previous chapter about using eBay Pulse to identify certain book selling trends and all that kind



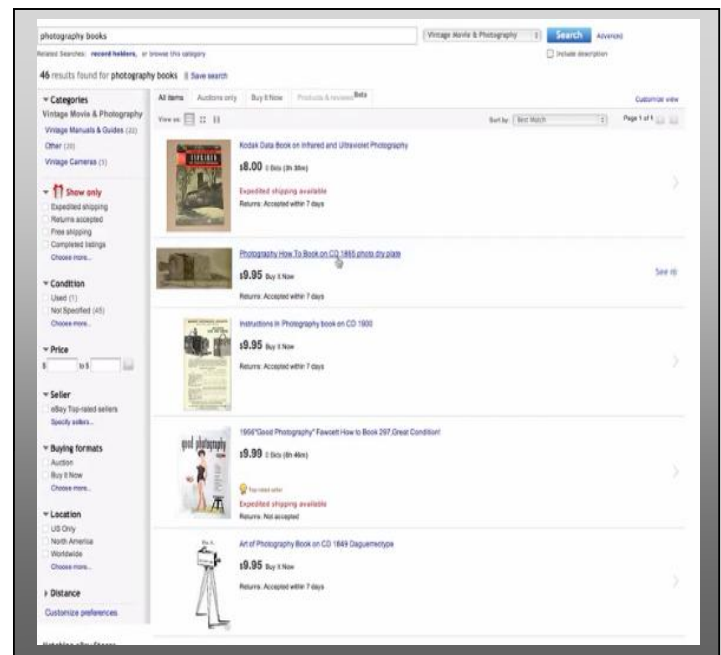


of stuff. But, what you may not realize is that eBay really is a good resource for finding books that are potentially in the public domain. Just like Abe Books you would want to verify the copyright status before you buy it but one of the things that is really cool is that eBay gives you the ability to track books on specific topics you are interested in. You can see here that I recently did a search on photography books so let's use that as an example. These are photography books, a lot of them are fairly new and I'm not interested in new, I'm interested in older books. You'll notice here that one of the options I have here is vintage movie and photography, which vintage would

normally applied to older books. If I click on that you will see that is definitely the case. This is a CD of somebody that put a photography book that was in the public domain on a CD. Actually a couple of these are. But I can also find vintage books on photography.

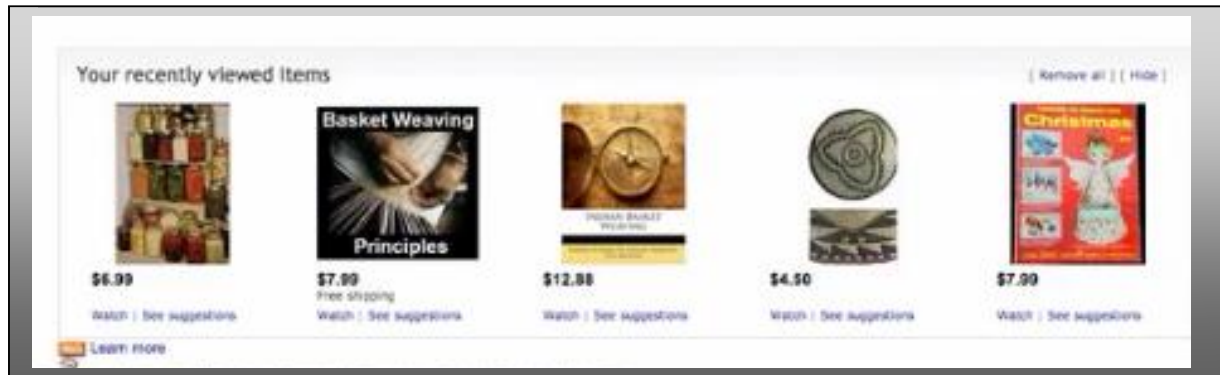
Let's say I want to track vintage books related to photography without having to

come to eBay all the time, wouldn't that be kind of cool if there was a specific topic or a specific author you were interested in, let's say I was trying to get a hold of a vintage copy of Napoleon Hill's Think and Grow Rich. Rather than having to come back to eBay all the time there is a much easier way of paying attention to what's happening on eBay without actually going there and that is through using RSS feeds.





We have this search result for photography books and we have limited it to just vintage, vintage movie and photography it has this little orange button that says RSS.



We want to click on that and that gives us an RSS feed for that particular focused search. If I copy that feed line, highlighted and copy it, I use Google reader, it's very straightforward and easy to use and I have it up in here.

All you have to do is click on the subscribe button, paste in the feed line that we just copy and click add. And there is our search results. Now, what is cool

about this is as people post new books or they bid on some of these



books that are in this particular search category this feed will be updated automatically. I can set up my Gmail account and all that kind of stuff to get notifications, if one of the beautiful things about Google's various tools is that it makes it real easy to track stuff so I don't have to worry about doing new searches on eBay, I can just set up a series of feeds and have them all come to

the Google Reader and track everything simultaneously, which makes my life a lot easier and a lot faster. I like that, and I'm sure you do as well. So that is eBay.

# Finding What You Don't Know To Look For

I have another resource that I want to show you real quick before we get into some of the how-to stuff here; it really goes along with the statement I make that

'you can't find what you don't know to look for.' I'm going to show you how to find what you don't know to look for.

*Find what you don't know to look for with Tony...Video 2a, beginning at 26:30*

You might remember that I said there are approximately 85,000,000+ books in the public domain and that number actually comes from an audit that the copyright office did themselves. That is not just some arbitrary number that I pulled out of the sky for the sake of selling books, it is a real legitimate number. So if you are unaware of how the whole copyright process works, when ever somebody publishes a new book, they submit to the copyright office and it is entered into a document called the 'catalog of copyright entries.'

Essentially what the 'catalog of copyright entries' is, it is a listing of every book that was submitted to the copyright office for copyright protection. The majority of those were never renewed but there are some that were, so there is a second document that the copyright office produces called the 'catalog of copyright renewals.' This is different because it only includes those books that were renewed, which we need to access that information to make sure the publication is in the public domain.

Think about this, every book that was ever submitted for copyright would be in these publications called the 'catalog for copyright entries' which in effect represents every book out there, all of those 85 million books in the public domain would be listed there. Wouldn't it be cool if there was a resource where you can actually search through all the entries, every book that was ever submitted for copyright and see what you could find? It would help make finding what you don't know to look for a whole lot easier wouldn't it? You could search by keyword, or maybe you have a favorite author and you don't know if they've published other books that you don't know about, it would be really cool to do that. Up until recently it just wasn't possible, as matter of fact you couldn't do it. There was no collection of those catalogs of catalog entries available anywhere. As a matter of fact, the library of Congress didn't even have a complete set. It actually took the collaborative effort of multiple universities together with Google to compile a collection of those catalogs of copyright entries. What Google did was, once they got all of those... Actually they started before they had the complete set, they just kept adding to it... they scanned them and ran them through OCR and created a special search engine to search through that set of catalogs that most people don't even know exists because there is no directly to it, it's kind of hidden. But I'm going to show you where it is.

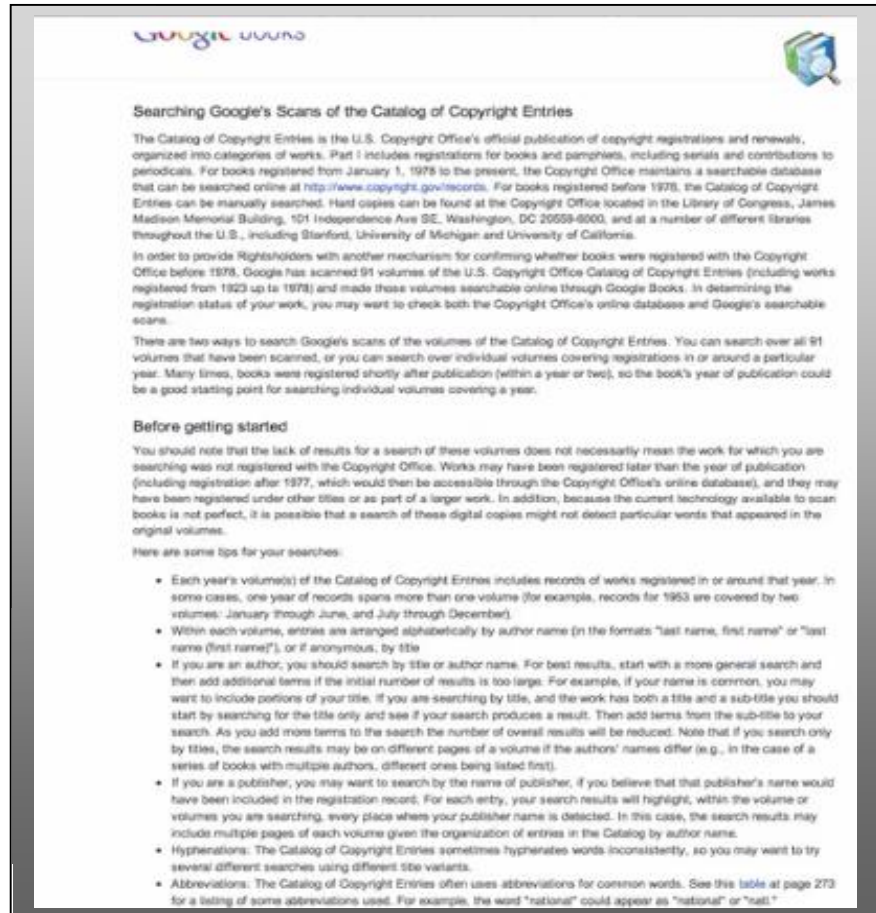
# Catalog of Copyright Entries

To search through the entire catalog of copyright entries, go to...

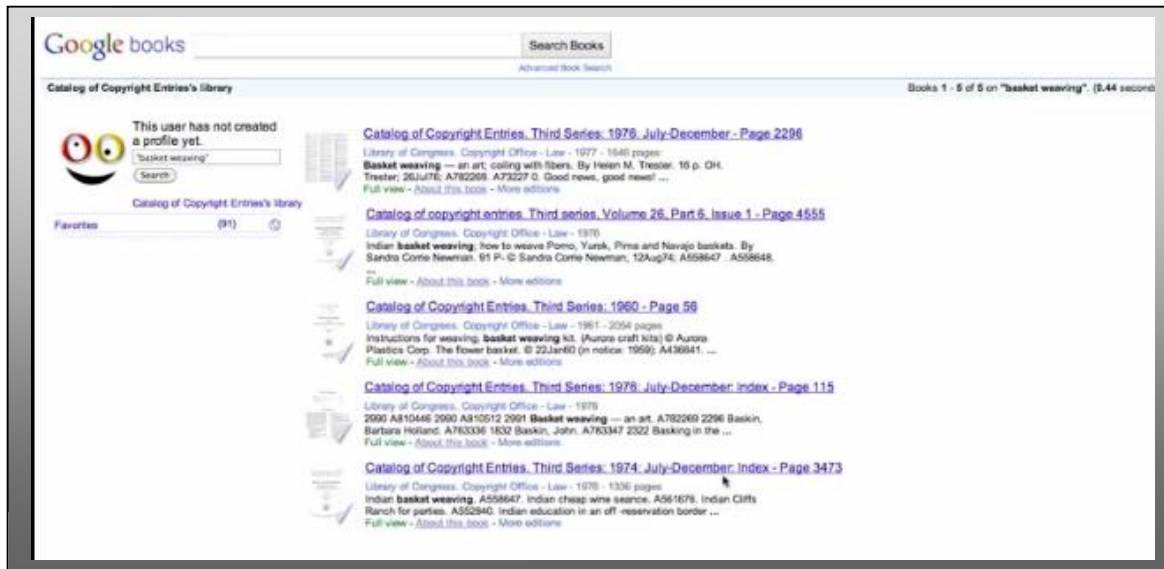
<http://books.google.com/googlebooks/copyrightsearch.html>. We are going to visit it so I can show you kind of how it works.

Whenever you go to that link, this is what it looks like and it explains a little bit about what I told you about. It gets into some search suggestions and all that stuff. They give you the ability to search the entire library from this search box located in the middle of the page. Or if you have an idea that there is a specific

book or author published in certain years you can actually scroll down towards the bottom and target specific year ranges. You can see where they have them all listed here.



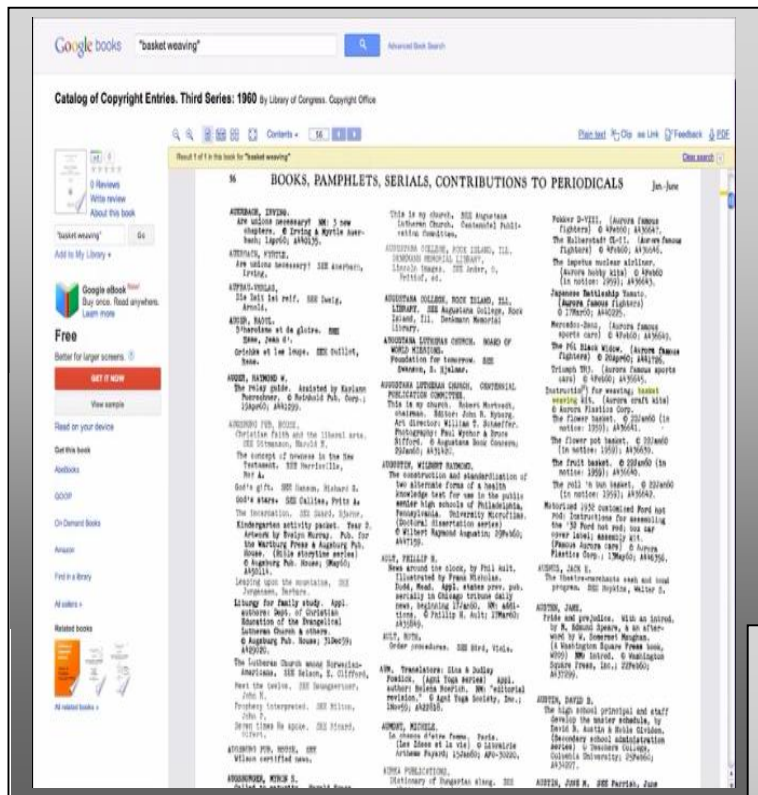
To stay true to form with our keyword, let's try basket weaving and I'm going to actually put it in quotes so that I know those two words are in the title. I haven't done this by the way, so I have no idea what we'll discover. Type in 'basket weaving' and click 'search.'



We have a number of different results, some of the results you were going to want to pay attention to and that is why I'm guiding you here are not going to be relevant to us because we know that the copyright laws changed in 1964 so any books published later than 1964, we can't use. So you can see here we have basket weaving but it is listed in 1977, another one listed in 1976. Here's one for 1961, that's a possibility...Instructions for Weaving, Basket Weaving Kit, that is actually a kit, not a book, but we can take a look at it if we want to. As a matter of fact, let's do that.

This is the actual catalog page and these are instructions for the kit, not the kit itself but it gives you that information. If they wanted to protect that by copyright it would have to have been renewed 28 years after this, in 1978 technically because it was copyrighted January 22, 1960.

That is one of the things I wanted to point out here, it gives you a lot of useful data.

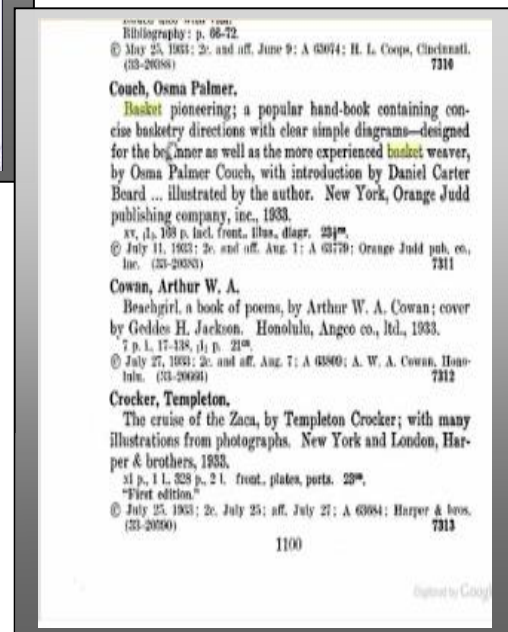


Let's limit our results and redo our search. It is only going to search it through 1960 so let's go back. What we are looking for are books that are related to baskets or basket weaving that are pre-1964. You can see that we have a lot of results, about 80. Some of

these are basketball stories which won't really help us.

One of the things I want to point out about using this...just because you find it here doesn't mean you will find a copy of the book.

The primary benefit of searching through this particular option is that it helps us identify a book we didn't know existed. What you do with that information, like this one, Basket Pioneering by Osma Palmer Couch, popular handbook containing concise basketry directions with clear simple diagrams designed for the beginning or experienced basket weavers. It was copyrighted in July 11, 1933 so that means, if it was renewed and we can check, it would have had to been renewed 28 years later which

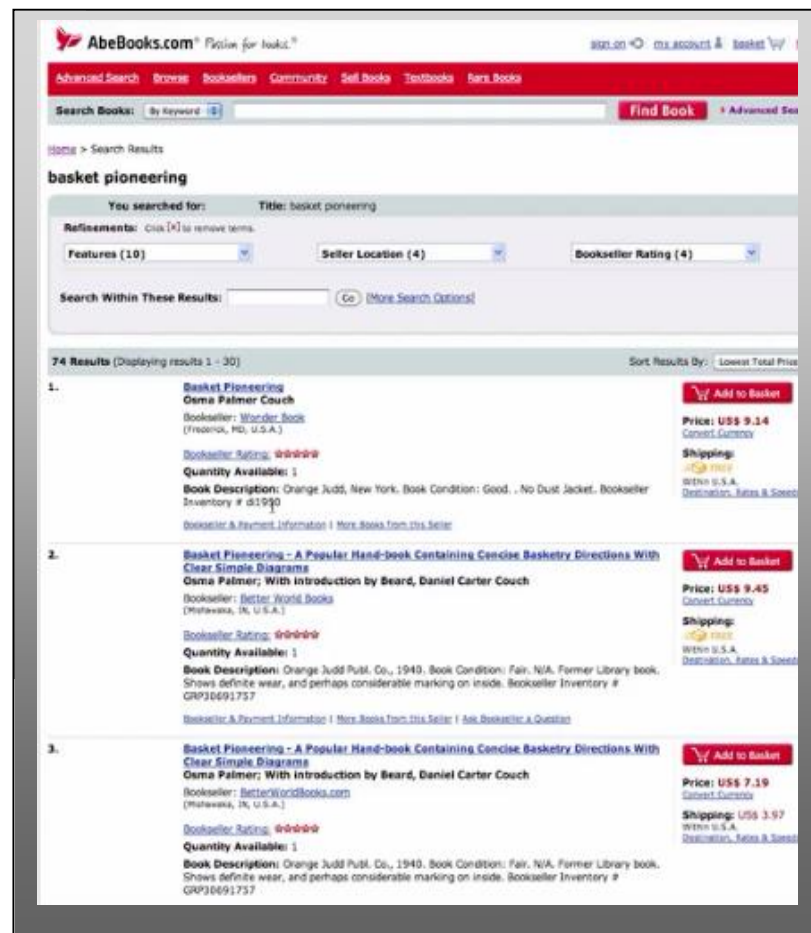




would be 1961 they would have had to renew the copyright. So let's say we check it using the methods in the previous chapter and we discover that it is in the public domain. Now the next step we would want to do is locate that book because it looks like it would be a pretty good book for us.

What we would do is take the title Basket Pioneering, come over here to Abe Books and see if we could find it.

And here it is, this is a 1950s version and a 1940 version so there are multiple versions available. Former library book shows definite wear. You can sort through to see which ones might be the best choice and then you can buy it for nine bucks. You can also do the exact same search on eBay if you so desire.



At eBay, type in Basket Pioneering and it is not available on eBay at all. The good news is that there are multiple copies of it available through Abe Books and that's a beautiful thing. You can pretty much choose which version you want, it looks like there's multiple published years so you want to verify the copyright status but you can get your own copy, scan it and then publish your own book whether to Kindle or Amazon directly, however you see fit. You can

turn it into a course or whatever. That is essentially how to use that research that we were doing with the catalog of copyright entries. It is pretty straightforward, that's how it works and I've gone through this method myself quite a few times. It works extremely well.

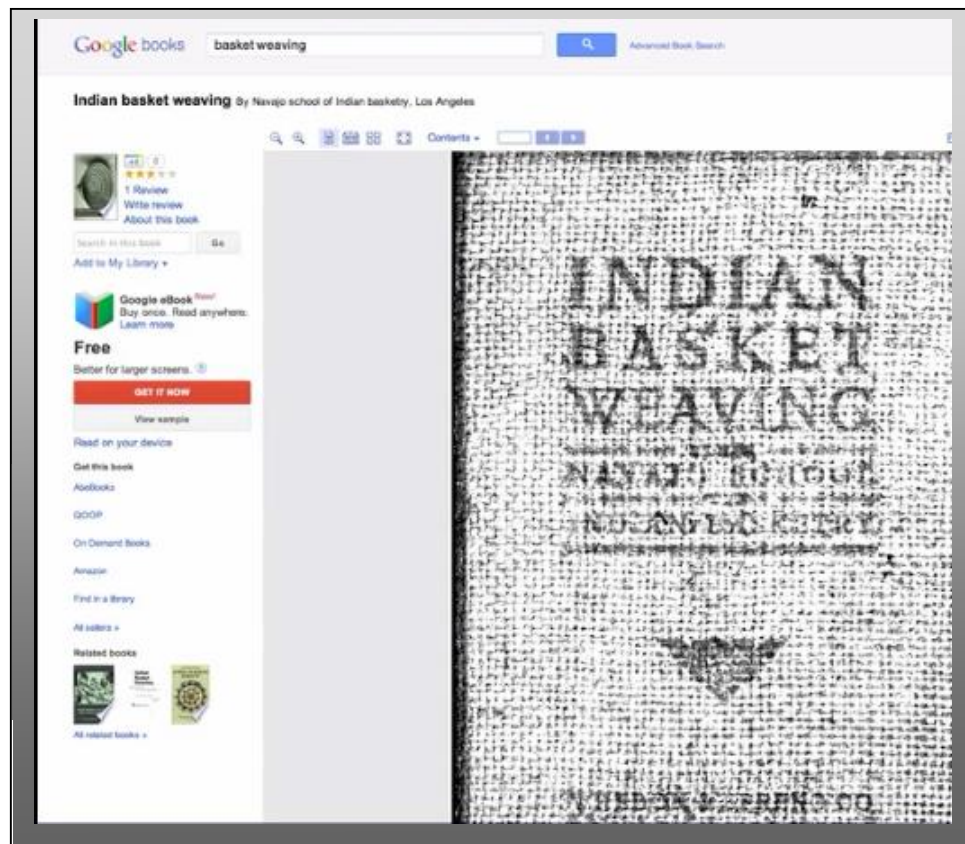
## Obtaining Editable Text

*Watch Tony as he demonstrates this process by viewing video 2a beginning at 38:54*

So, let's move on to the next which is obtaining the editable text. As I indicated when we were on Google, getting editable text is not always the easiest thing

to do. In most cases as we saw on Google books and on Hathi Trust, they give us the ability to download the PDF and we can download and convert it using the

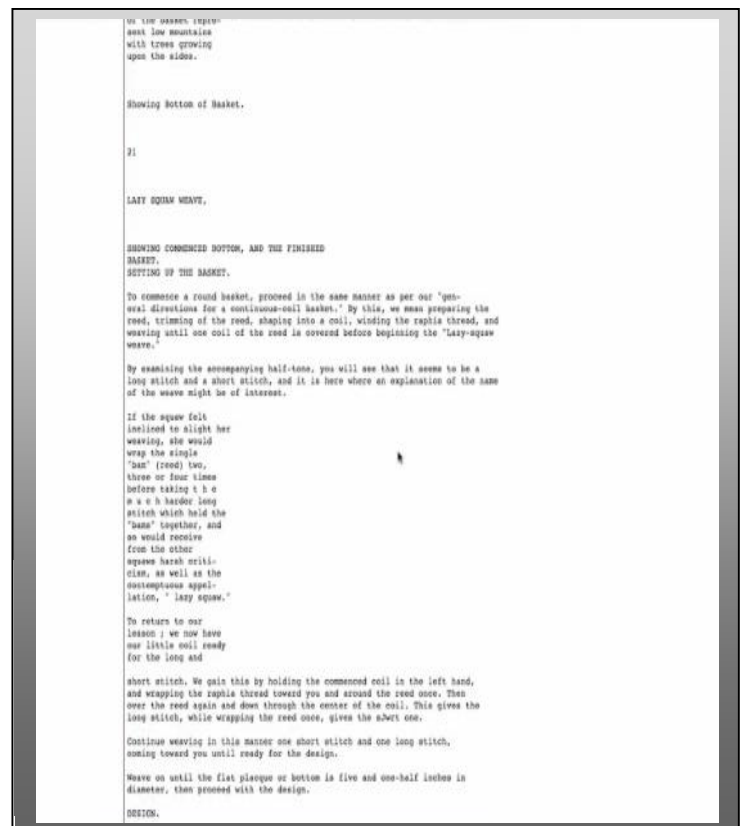
OCR software. I'll show you what I use in a little bit or we can copy and paste



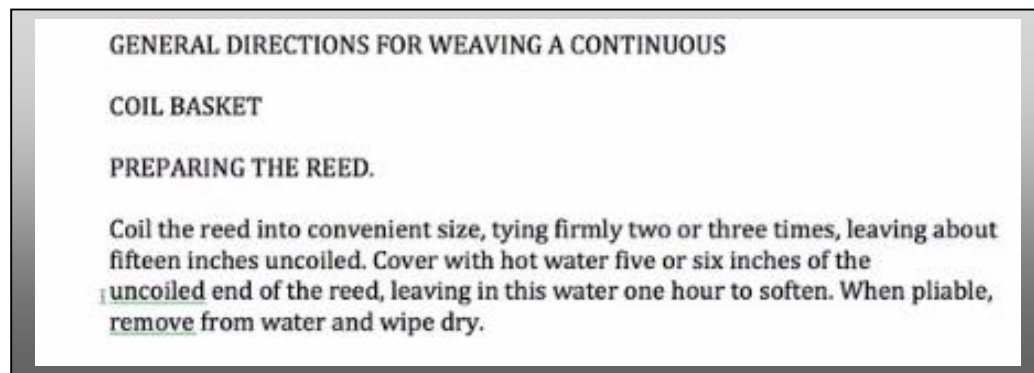
the text parts from Google and Hathi Trust. With Archive.org, they give us the ability to download the full text.

You can see from our example, here is our original book, Indian Basket Weaving that we found on Google and you can see here it is just the grayscale images are not high a quality, they are yucky looking and all of that. We can download the PDF by clicking on it, which I've already done that. On the archive.org site we have completely different version and we can download that PDF as well but, in this particular case and not all books that are on Google are on archive.org, just so you know, it's not an exact duplication but there are quite a few. In this case we can download the full text if we wanted just by clicking 'full text' and you can see it all right here.

The gobbledygook you have here is where the images are, by the way. Now you might also notice that all the hard returns are in here and I'll show you a fun trick I use to remove those, because that is all part of the editing process...so, the only problem here is that we can't really tell for sure where the images go if we wanted to include those images. And as we saw with the Google book, the PDF, we've a problem anyways because the images weren't that high quality and they certainly won't reproduce well in print form. In digital form perhaps, but not print so if we wanted to make a print book



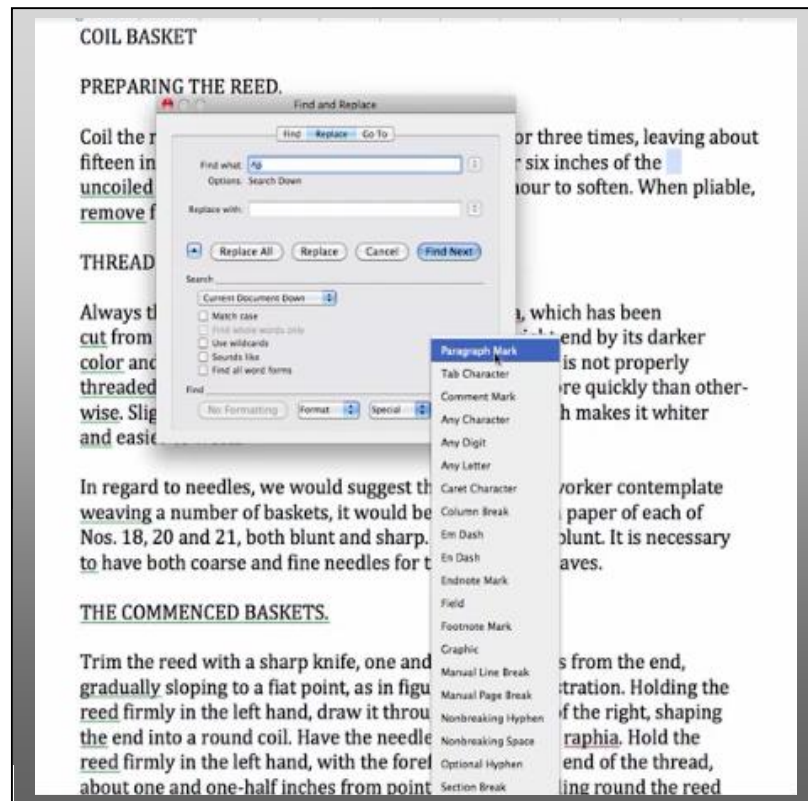
out of this, we'd be in trouble. So, if we want to pull the full text from archive.org it is simply a matter of highlighting it all, scroll down to the end, hold the shift key down and click and that would allow us to copy the entire block of text. Let's do that, I'll hit control c, go over to Microsoft Word and hit paste, it will paste everything in there. Normally I don't do that, I don't do a direct paste into Microsoft Word. In this particular case it may not make a difference, but what I normally do is I actually paste 'special' because that gives us the option of how it pastes in. So if there's any formatting from the website the normal paste



automatically paste using formatted text so it would retain any type of formatting in there which could cause some formatting problems for you. What I normally do is choose 'unformatted text' and then click 'ok.' It pastes it and strips all of the formatting out of it. So you can see here, the font and everything looks completely different because it stripped out all of that formatting.

We still have a problem and it is these hard returns. If I hit 'delete' it is going to pull the word about up and you notice all the green underline here, that means there is a hard return that exists before this word. We want to remove those because it affects how the text flows in our final book, if we are creating a print book or whatever. So you might think, holy cow, I can't go in and delete all of those spaces throughout the entire book but there's actually a very fast way of doing that, the fast way is...we are going to pick an area like right here and we can see there is a hard return there because we have the

green underline. In Microsoft Word, we are going to highlight this space; we don't want the word 'the' so we will highlight that space right there and copy it and then if I hit control f which is find and replace, and paste you'll see I get this little carrot with a p. That actually is Microsoft Word symbol for hard return, for paragraph break.



Another way to discover that is to come down here and do a search for 'paragraph marks' or if you want to do a search for anything else, which in this case what we want and if that paragraph break. Instead of copying like I just showed you, you can come down here and choose 'paragraph mark.' So, we can identify the paragraph mark which is the hidden character and we don't want to replace it with anything, we just want to remove it. We have one highlighted so if we click 'replace' notice that in our text in the background, it automatically removes it and moved onto the next one. Then it becomes a matter of replace, replace. In this case we don't want to remove the returns here because this is a subtitle or subheading so we would just click 'find next', 'find next', 'find next' and then replace, replace, replace. Notice here where in the original text it was actually hyphenated so we would want to go in and remove that hyphenation, we don't need that.



Now you might think this will still take a little time, and it will, there's no doubt about it. But, the good news is, it is a lot faster than just going in and using the delete key, here you are just clicking a button. This is part of the text cleanup process. Once you get into the routine, it you can clean it up really fast. For instance, we can go in and highlight that whole section here and delete the whole thing, so all of that is removed. Then we can go back in and out all the images if we want to, we don't have to but if we want to we can add those images in. You might remember that we have a problem, we need high-quality images and we don't get that from Google books. That's where archive.org comes in handy, if you use it the right way, the way I'm going to show you.

## Creating Better

## Images

Let's go back over here to archive.org and you notice that we have multiple choices, we have read online, PDF, black and

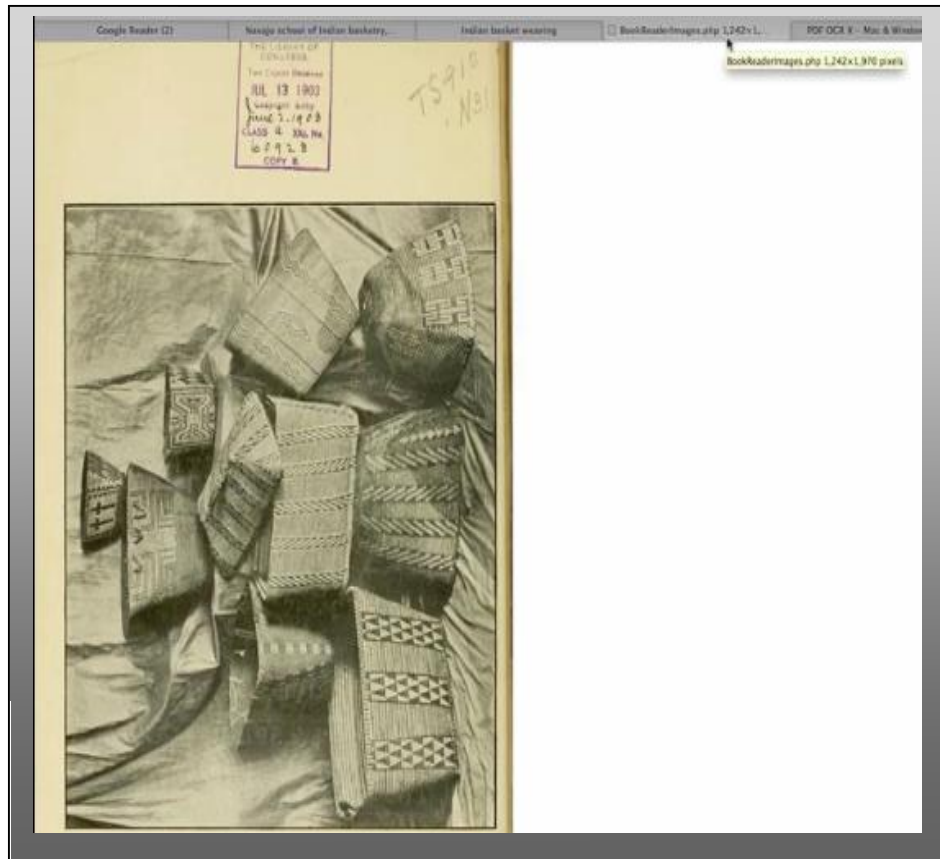
white PDF, so on and so forth. To get the best possible images out of books that contain images, the version that you want to focus on is the 'read online' version and I'm going to show you why. If you click read on why you notice it gives you the book and two-page spread and you might be thinking, wait a minute I can't edit this or anything like that and whenever I click on it, it just turns the page. And that's true. I, here is what I want you to do...if you click on single page you have your illustration. If you right mouse click on this page and choose 'open image in a new tab', I want to show you something because normally you would think you would just save the image and that's good. Except that is not exactly what you want to do and I'll show you why. If you click 'open image in new tab' and switch over here to our new tab, you'll notice you can zoom in

*To watch Tony  
demonstrate this process,  
go to video 2a and begin  
watching at 53:54*



and it gives us a half decent looking image but up here in the tab, notice this right here, it shows us the dimensions of our image, 1242 x 1970, that's the size of this page which isn't too bad but, let me show you something.

If we come down here and start clicking on the 'zoom in' tab one of the technologies that they use on this site is scalable pages so if we zoom in to the larger size, like



really big and then right mouse click and open in a new tab I want you to see something, now our dimensions are 2483 x 3940, so just by us zooming in and then changing it, there comes a point in all of this is me in there are only actually two sizes of this page, the 1242 and the 2483 which by the way if you haven't figured it out is essentially double. In other words this is like medium and this is large and large is double the size of medium. They look the same except in this one when I click on the zoom it is that size, and this one when I click on the zoom it is much bigger which gives us a higher quality to work with. Now you can see where there's printer dots in there and we can use a simple de-blur or blur in Photoshop to remove those. My point is whenever you are pulling images out of the book you want to zoom in on the 'read online version' on archive.org and

then save the image and that will give you the larger version that we want, in this case 2483 x 3940.

In some cases you will be blown away, that experience to birds, flowers and animals, all that kind of stuff, some of the books that they have in those genres in archive.org you can get some amazing images at nice size resolutions just using the method that I showed you. I don't know if anyone else is even teaching this but this is key if you are interested in quality and quality is something that is very important to me, that is why I use this approach.

Now, let's say that we find a book that we'd like on Google and we can't find it on archive.org and we don't feel like copy and pasting all of the text parts into the word document because that will take too long and we're not feeling like taking that much action. We have another option, and that option is to convert that PDF into editable text. There are a number of PDF to text converters out there but not all of them read images.

What I mean by that, the PDFs you download from both Hathi Trust and Google are images that have text in them which means you need an OCR engine within your software to recognize the text and make it editable.

Over the last few years I've tested a number of different

**PDF OCR X**

Getting Started | Language Packs | Support

**PDF to Text**

- Simple drag-and-drop utility for Mac OS X and Windows
- Supports over 20 languages.. [Learn More...](#)
- Also works with images such as GIF, JPEG, TIF, BMP, PSD, PNG, and more...
- Converts to text or searchable PDF (New in version 1.9)

**Convert Scanned and Image PDFs into Text Documents**

PDF OCR is a simple drag-and-drop utility for Mac OS X and Windows, that converts your PDFs and images into text documents or searchable PDF files. It uses advanced OCR (optical character recognition) technology to extract the text of the PDF even if that text is contained in an image. This is particularly useful for dealing with PDFs that were created via a Scan-to-PDF function in a scanner or photo copier.

**Requirements**

- Intel Mac running Mac OS X 10.5 or higher
- Windows XP/Vista/7 with Java 1.6 or higher installed

**DOWNLOAD NOW**

**FREE Community Version**  
Limited to single-page images and PDFs

- [1.9.14 for Mac OS X](#)
- [1.9.6 for Windows](#)

**\$29.99 USD Enterprise Version**  
No limit on the size of the PDF

- [1.9.14 for Mac OS X](#)
- [1.9.6 for Windows](#)

**How It Works**

- Drag the PDF that you want to convert onto PDF OCR X.
- Select your conversion settings (e.g. language, output format, etc...)
- PDF OCR X converts your document to text or searchable PDF.

**Installation**

- Download PDF OCR and place it in your Applications folder.
- Drag the PDF OCR icon down to your dock so that you can easily access it.
- Now you can drag your PDFs onto the PDF OCR icon on your dock to have it converted to text.

**Features**

- Works with any PDF, whether it is a scanned PDF, or a PDF generated from a document.
- Easy drag and drop interface.
- Support for multi-column documents and advanced formatting (New in version 1.3)

**Disclaimer**

PDF OCR uses OCR (optical character recognition) to convert images of text into text. While the technology is in the field of distribution, neither

software programs and there are a few out there that work reasonably well but I like to look for the ones that don't cost a huge amount of money but gets the job done. There's actually a program I just started working with that is available on both the Mac and the PC, called PDF-OCR-X (<http://www.solutions.weblite.ca/pdfocrx/>) and it cost about \$25 or \$29, not really expensive. One of the things that I like about it is that it is really simple and straightforward to use, as a matter of fact I will show you how simple and straightforward it is.

It works pretty much the same way on Windows if you are a Windows-based person. It says 'drop or paste your PDF or image here.' You can also have JPGs by the way's that have text in them and you can drop them in here as well and it will recognize it. I downloaded the Google PDF of the Indian Basket Weaving document that we found so I will drag it into here. It will take a little while to convert it into text. It brings up these settings, it's one file of 121 pages and we wanted to it to output to text, not a searchable PDF, and we want that content in a single column. We have a couple of different text wrap options, we have soft wrap and hard wrap. It in this case we want soft wrap, the difference basically is hard returns versus software and. If we do soft returns we don't have to go in and remove the hard returns later. Then we just click 'convert.' What it will do is begin reading the images that are contained within the PDF, it processes those using the OCR engine that is built into the software and you can see that it actually processes through pretty fast. It's doing about a page a second, so 121 pages you are talking about 2 min. which isn't bad at all. The previous software that I used was much slower; it was like you set it and then went to bed. This one is actually much faster. Rather than sitting here and watch it do it's thing, I actually already ran this so I will close it and this is what we ended up with.

We can see here where this is the Google stuff at the beginning , we don't need any of that. This is just the text file, I can copy and paste it into Microsoft Word very easily. I just realized what it was scanning actually replaced the original document that I had so this is the number of pages it converted while we were waiting because this is like the first seven pages or so. Now, I can take this and paste it into Word, highlight all this other stuff and because it's text I don't need to paste special, there's no formatting in it. So let's delete the other one and paste it in there. You can see where it retains paragraph breaks but if you want to add a space between them it makes it that much easier, you can highlight this, choose paragraph, space before, 10 point and there is a table of contents so we might want to go in and edit that.

Any OCR you work with is never perfect, there's always going to need to be some kind of editing. You can see here where there was a hard return in the original book for weaving so we would need to go in and fix that. Here is a period rather than a comma so you would need to fix that, this single quote is wrong.

(AUDIO IS FUNKY...57:58)

They have a funky way of spelling fibers which we would want to correct so this is the edit part of the process that you will need to go through. But all in all it didn't do a bad job. On page eight or nine, whenever I stopped it but the good news is, since I stopped it early we still have access to pages that are done. As you are preparing your text, you can go back and access your original PDF.....one of the thing you want to be aware of is you want to make sure you don't leave the existing page numbers, you don't want to leave them in here. The other thing you need to watch for is some cases are the title pages ????????

so you would want to delete those too. So, it is a matter of going through and doing that final formatting.

We can take the image, we copied it from Archive. Let me show you this, let's save image to downloads.

# Editing Images in Photoshop

What I'm going to do is open up Photoshop and show you what to do with the images.

We'll let Photoshop load here will quick, I

wasn't planning on doing this but I think it will be real helpful to you. Ok, so we want to open the image and then crop off all of the extra, we don't want things like the black border. The next thing we are going to do, because of the way it was published in the book at an angle, lets rotate that bad boy, so 'image rotate 90 degree clockwise' and if we zoom in, which control + or command + in Photoshop will allow us to that, you'll see that we still have the dots from the printed image. There are a couple ways we can remove those.

*To watch Tony demonstrate this method, go to video 2a beginning at 59:56*

The one that I normally like to use, we could use 'high pass' will work, we can use 'median or de-speckle'. Let's try de-speckle and



you'll see that it removed it but not a lot. So let's go back here again and try median. Median does a much better job so if we increase the radius at a 2, you can see where 'see the original' like if I turn the preview off, that is what it looks like before median if I turn the preview on that is what it looks like after. You'll notice that it is a little blurry, in fact that is what we're doing. You can actually go in and sharpen it later and that will help, so let's apply that and then we can apply a 'sharpen' filter and that will help sharpen it up a little bit.

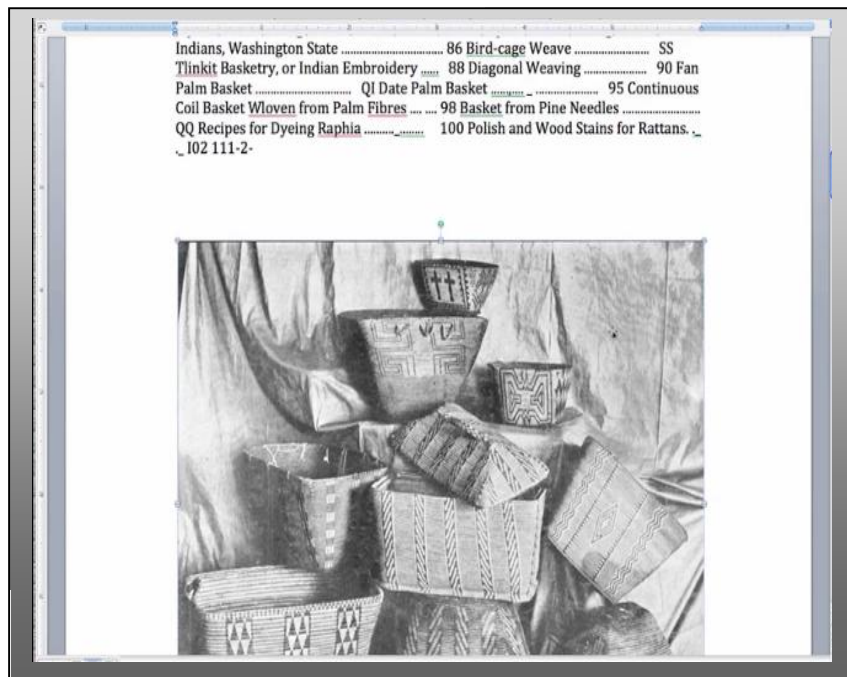
Now, the image is much more usable. But it still has that cream cast to it and that isn't good. Let's zoom back out. The next thing we want to do is make this pure black and white. We could very easily come up to 'image mode' and choose 'grayscale.' What that will do everything that had that tan color from the paper, it's turning it gray and that isn't necessarily a good thing either. We need to fix that. The easiest way to do that is use 'levels' in Photoshop. Command or Control L actually brings up levels and levels have this cool feature where you can control the shadow highlight. In this case we can't do the mid-tones so we want to actually sample the highlight and we can control what that is going to look like, which is this case, anywhere we click our dropper, it will go 100% white.

When the book was originally printed on white paper; that is what it would have looked like so we are going to leave it as it is. Watch what happens, with that highlight eye dropper selected, I come over here to pick an area where it is pretty much the pure gray, where the original paper would have been. When I click on that, it instantly re-sets that to white. So you can see it didn't affect anything else in the image, just the original tone. Then, we can adjust the image; we can add a little more contrast and darken it if we want, whatever the case may be. Maybe the white is too much so we can go in, double click and change to perhaps 5%. Now, if I click on the white area, it actually changes it to 5%, see what it did? There's our original and there is our current

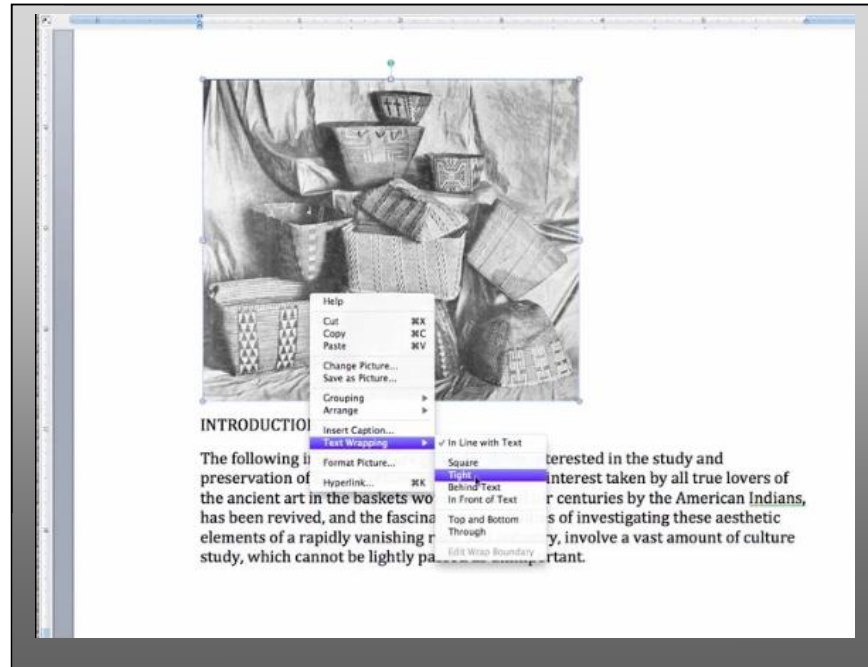


version, which 5% makes more sense than 100% white. Click ok, we'll save that as a default, so now we have our image and all the little patterns are gone, it is fully grayscale and we can just save it as a tiff or jpg image and then import it into our book.

Let's 'save as' and then come over to Microsoft Word and let's say we want to add it before the introduction. You want to insert a picture from a file and we want to find it on our desktop, which means that we want image...insert...and there you go.



Now, the image by default in Microsoft Word is imported in line with the text, meaning that the text won't wrap around it, it is in line with it. You can actually change that if you want to, say, reduce the size of the image. If I reduce the size of the image here right now, it will bring the text up but the text can't wrap around it. Let's say we want the text to wrap around it so we can come here, right mouse click, choose 'text wrapping.' What I normal use is 'tight', you can see here where it says 'inline with text', if we switch it to 'tight', look what happens.



You can delete that extra space and now the space wraps around our image. So you can see how using these methods I'm showing you, you can rebuild a book really in any format that you want. The original book was probably 6x9 and right now we're working at 8.5x11.

Now, I know that I went through a lot of this pretty fast, but that's the beauty of replays, you can watch the videos over and over, go back and see exactly what I did and how I did it. Now, this also assumes that you have tools like Microsoft Word and Adobe Photoshop. But, Photoshop is extremely cheap, Adobe now has a subscription service where you can actually subscribe to Photoshop on a month to month basis and it is very reasonable. So the tools aren't out of anyone's reach.

## Book Checklist

Whenever you are assembling your book, you go through and do all that I showed you and you are getting ready to either create it for Kindle or as a print book, there's a couple of things you always want to keep in mind. First of all you need a cover, whether you create your own or outsource it, you want to have some form of cover and that would normally include the title and the author, whether that is you or the original author, edited by you or however you choose to handle that. You want to have some form of a copyright page because remember, we're adding value just like I said in the last video, we're adding value and that gives you the ability to copyright your product. You want some sort of a table of contents so people can find their way through your book. The text for the book itself. You want to also make sure you include your contact information and that could be 'about you', your phone number, email, mailing address, website for sure. And then also ads for other products, whether they are your products or affiliate products, you want to have links to your website or to other products available on Amazon. Or to videos you have on You Tube. There are a number of ways you can spin this. If you really enjoy Indian Basket Weaving, make sure you check out my free video series that I have on You Tube, here is the address to that channel. That way you are using your book to drive traffic to your other products which is what we want. It helps build your credibility and helps build the relationship you are building with the customer and it gives you the opportunity to make more money, which is a good thing.

## Homework Assignment

The next chapter is about magazines and your homework assignment is find one book...not 10 or 50, because I know it is tempting to go on a crazy mad rush, not sleep at all....find ONE book, whether it is on Google, Hathi Trust or Archive.org. Using the formatting and conversion methods I showed you, I want you to convert it and add unique content, whether it is 'points to ponder' bullet

points at the end of each chapter, whether it is a questions section, whether you mash multiple books together, come up with something completely unique. I don't care what you do, but I want you to add unique content and I want you to format it as we discussed and how I showed you in Microsoft Word. You can do the exact same thing Open Office, which is free so you don't have to use Microsoft Word if you don't have it. The important thing I want you to do here is to actually do it. It doesn't have to have images if you don't want.

If you want some extra credit from me...and I don't know what that means...upload it to Kindle. Some of the bonus training I am adding will take you through that process so you will be able to do that.

That is your homework for this week. The whole purpose behind it is to get you to take action. Why invest in this course, even if it is a great value, if you aren't going to do anything with it. Really.