



BOOK INDEXING BASICS: LEVERAGING THE POWER OF INFORMATION

OPERATOR: Good afternoon and thank you for standing by. At this time, all participants are in a listen-only mode. After the presentation, we will conduct a question and answer session. To ask a question at that time, please press star, one. Today's conference is being recorded. If you have any objections, you may disconnect at this time.

Now I will turn the meeting over to Mr. Chris Kenneally. Sir, you may begin.

KENNEALLY: Thank you, operator, and welcome to everyone on the call. My name is Chris Kenneally. I'm the host of the Beyond the Book programs sponsored and presented by Copyright Clearance Center. I want to welcome everyone today to a very special program we call "Book Indexing Basics: Leveraging the Power of Information."

Joining us today is Seth Maislin. Welcome, Seth.

MAISLIN: Hello. Glad to be here.

KENNEALLY: Seth is the immediate past president of the American Society of Indexers, and he's going to address for us a range of issues, including the basics of good indexing, the first steps towards writing your own index, and also – this is important for publishers but also for authors – how to judge index quality.

It seems to us here at Copyright Clearance Center that authors and publishers will agree if you want to really enhance the value of your works in the marketplace, a good index provides a far better understanding for how the content is included in the work for the audience, and that's going to be important to you. Indeed, we think that a well-made index can help you leverage better the power of the information age for your own business success.

Let me tell you a bit about how the program will work and we'll just dive right in. Some rules of the road. We appreciate your commitment of the next hour to us, and in return we will promise you that it will last an hour. If we say it's an hour program, that's what we're going to make it. We will leave some time for questions. You'll be able to raise your hand, electronically speaking at least, by pressing star, one on your phone. But if you would like to e-mail me a question,

you are welcome to do so. I have my e-mail open in front of me and that address – pretty simple enough – chrisk – C-H-R-I-S-K – @copyright – C-O-P-Y-R-I-G-H-T – .com.

As the operator mentioned, we are recording the program. We will be presenting this in our podcast series later this fall. It'll be in a two-part, half-hour each program, so at some point about halfway in, you'll hear me kind of wrap up and come right back, so don't go away. It's just for that purpose.

The other thing we want to tell you about is please do go and look at www.beyondthebook.com to learn all about our educational program. We try to connect authors, publishers, and others in the information content industry with experts who can discuss all the latest business issues from initial research to final publication and beyond. And I think in this particular case, we really mean the back of beyond, and that's the index, of course. We appear sometimes on CSPAN, quite frequently on Book Television in Canada, and you can always get our programs again at beyondthebook.com.

So let's introduce Seth and get this started here. Seth is a managing partner at the indexing firm Potomac Indexing. He's an adjunct instructor at three Massachusetts colleges. That's kind of like a hat trick, I suppose, of teaching. And he is an experienced trainer. He's one of the foremost indexing and information architecture consultants in the country. No surprise then, as I mentioned, that he is the immediate past president of the American Society of Indexers. If you want to learn about that group, you can go to asindexing.org. That's asindexing.org. In his own 15-year career, Seth has developed a wide range of specialties including teaching indexing and related topics as well as indexing computer, technical, and general interest topics. He's the author of a website and blog on indexing and information architecture, and that is at taxonomist.tripod.com. And if you don't get any of that, please to e-mail me. We'll be sure to get you all the details on all those websites.

And let me share with you just a real quick thought. I've been an author myself and at one point in my own career discovered that I had signed a contract that required me to do an index for my book. I hadn't realized that until the editor reminded me. Many people may ask, why care about indexing? And I suppose my first answer to that question is if you care about books, if you care about creating information that's of value to readers, you care about everything that goes into publishing. You care about the look of the book, the feel of the book. You care about the font and you care about the design, and really that includes as well caring about the index in the back of the book. After all, before there were search engines, there were indexes and fingers to go through them.

So Seth, I wonder if we can start with the theory and the history of indexing to kind of put this in some perspective for people. Tell us about that.

MAISLIN: Sure thing. Well, there are so many ways to sort of talk about this. The basic bottom line of what an index offers compared to something that has no index, that an index provides multiple ways into the content, ways that wouldn't have been available in a book that doesn't have an index. Think about – imagine trying to just find something somewhere in the book that you read, and of course, the easiest example of that, books that have no indexes, are fiction. Although it doesn't really make sense to index a novel, imagine the last novel that you read and there was a passage that you particularly liked, a quote, or maybe you're reading about a character you don't remember being introduced to and you have to go flipping for it. And that's the experience of not having an index is you have a reader who wants to find something and there's really no good way to find it except flipping through. That's what an index offers, and that's true in basically any environment.

One of the stories that I like of what the world was like before the indexes as we know it goes back – I don't remember the year, but let's imagine that documents were being made on stone tablets. These are documents that deal with land ownership and financial constraints and taxes and so forth, and suppose there's a dispute and you need to go find this stone tablet. Well, the stone tablets were basically stacked human height in a big giant basement space. And if you had to find the stone tablet that had the information on it, how do you do that? Even if you knew exactly where it was, remember, you'd still have to lift off all the tablets that are above it in this stack, and how long that would take. And then you better make sure you put them back in the same order. So here's an environment where you need some external tool that's, say, not as heavy as a stone tablet, that at least points you in the right direction to find what you want.

This is the nature of indexing. It's been around for a very long time. As human beings, we like to remember where things are. If you've ever walked into somebody else's kitchen and been unable to find a spoon, well, there's the kind of situation where an index-like device would be useful. That's what librarians do to help you find books, what directions do in terms of a map to help you find where you're going. That's fundamentally what indexing is.

KENNEALLY: Seth, can I say how I relate to that. First of all, the idea of working with those stone tablets makes me think that indexing would be good exercise as well as good for the information there. But we all know if we walk into a bookstore with nonfiction, if there's a history book or a book on a subject that we deeply care about and we want to see how the author treats that subject, we will go straight to the index. We know that that's important. And we all know the feeling of disappointment if we find out that that book the history of Massachusetts doesn't have an entry for Samuel Adams or doesn't have an entry for Paul Revere. We would know right away that this was not something we wanted.

So the index is really almost immediately – it's funny. We think of them as being in the back of the book, but for many readers – researchers – it's the first experience they have with the work.

MAISLIN: Let me clarify something about being in the back of the book. The reason indexes are in the back of the book is pretty simple, and that's that they have page numbers and the index is the last thing that gets written, at least in the traditional publishing process. If the index could be written first, it could be put in the front, but until everything is on its pages and you know what those page numbers are, you can't start the index, and you're not going to write an index and then shove it into the middle of the book. So the reason it's in the back has a lot to do with the publishing process and not about the use of it.

In fact, I recently did some consulting work with Mercedes Benz. I had the pleasure of writing the index for their E Class, one their most – one of the vehicles they just released. They were thinking about taking the table of contents and not using it and putting the index in the front. And the reason they can do this in Germany is because for whatever reason, German publications generally did not and do not have indexes. It's a relatively new technology over there.

So they're in an experimental phase in a publishing world where it is now possible to throw something at the beginning of the book, in the Roman page numbers, for example, and not affect the pagination.

I've gone to a talk about technical writing, and the speaker opens the indexing talk very simply. He says, you're a writer. What's the last thing you do when you write? What's the last thing you write? And everybody says, the index. And then he says, OK, now you're looking at a book. What's the first page you turn to? It's the index. So the fact that it's in the back of the book has nothing to do with its value. In fact, it's extremely important.

KENNEALLY: Absolutely. And what you told me when we were preparing for this was an indexer needs to think about having the whole book in his or her head at one moment, and that's critical and it's a really important difference between how authors think about the book.

MAISLIN: Yes. The way you have to write a book – and this is from the author's perspective – is one of narrowing focus. You start by thinking about the whole book. You outline the whole book. It's almost as though you write the table of contents before you write the book. You have the big picture in your mind. And then you choose one chapter and you work on it. And when you do that chapter, you're going to start with one section of that chapter and one paragraph, and then the very sentence that you're writing, down to the word that you have to write next. And once that sentence is finished, you move on until the paragraph is finished, the section is finished, the chapter is finished, and then you move away. And in a way, if you're still thinking about that chapter as you move forward, it interferes with your ability to write. So being able to narrow your focus that way and look at one word at a time is the key to really brilliant writing.

Indexing, on the other hand, is exactly the opposite. Every word that you read you have to retain. You have to maintain the flow of ideas so that if an idea appears in Chapter 3 and Chapter 6 and Chapter 10 and in Appendix B, the indexer has to remember that in some fashion so that the index can have a single entry with four page numbers, one for each of those locations.

And also the relationships of all of those ideas. Something in Chapter 3 is in one context. Something in Chapter 10 might be in a different context. And so how you create those contexts and the way in might help someone going into an index distinguish between do I want to turn to this page number that's in Chapter 3 or do I want to turn to this page number that's in Chapter 10?

So in a way, the very map that the author starts with is what the indexer has to, in some extent, create during the indexing process, but at the same time not be limited to. Authors know that there are many different ways in which you can organize a book. All of those ways have to be implemented in an index because you don't know how the reader is thinking about it. You don't know what their sort of table contents, their mental map of the subject matter is when they go in. An indexer has to provide lots of different ways into the content.

This, by the way, is one of the fundamental advantages of an index over some of the other tools, so I know there tends to be confusion between, say, a concordance and an index. A concordance is a tool which is generally computer-generated where the words are extracted and lifted in alphabetical order. The earliest concordances were for the Bible, where you might be interested in saying, how many times does the word bread show up in the New Testament? I'd like to know that. And that's fine, but what concordances don't do are provide multiple ways in. They don't look at synonyms for bread. They don't look at context in which bread is used. It doesn't connect bread to, say, starvation or eating or bounty. No symbolism, no context.

So what the index does is it does provide all of those ways in. All the different ways people can think about things should be captured in the index, and the only way to do that is to have the whole thing in your head at one time.

KENNEALLY: It sounds like a daunting task, and I think, again thinking back to my experience being told that I was going to have to do an index, it terrified me, I'll confess, because I realized that I was going to have to think about just these things and all at once. It really must require tremendous concentration.

MAISLIN: I would love to say that it is the most daunting task in the world because that justifies my rates for doing my job. The reality is that indexing gets easier with experience. The more you index, the better you are. The same way if you do a lot of crossword puzzles, you get better at them. The same is true with indexing.

The more you know a subject matter, the more you already have a map in your head to work with. So I hate indexing economics books because I know very little about economics and I'm not particularly interested in economics topics. So I have a very poor table of contents for economics in my mind, and therefore, doing that index, I would do a pretty poor job in comparison to topics that I'm much more familiar or interested in.

And there are tools that you can use when you index to help you keep track of things because yes, for particularly large books – and the most dramatic examples are multi-volume encyclopedias, which are then often indexed by many, many people who have each their own sort of mental table of contents – you need tools for collaboration to do that.

But it is a daunting task and it is fundamentally different from the writing task. One of the challenges that authors face when they write their own index is they have to turn off that writing characteristic of narrowing focus. As you're indexing, it wouldn't surprise me, as an author, to say, oh, I could write this better. The reality is by keeping it all in your head, though, you will actually be a better writer. If you know that you wrote something in Chapter 3 that's related to what's in Chapter 10, that's extremely informative, and a lot of authors miss that. Probably my favorite author feedback for an index that I've ever heard is along the lines of, wow, I didn't even know I wrote that. That's the perspective that indexing can provide.

KENNEALLY: Right. It's providing some kind of connective material for the entire book. It's a road or at least the TripTik, if you will, through the book, and I'm thinking again on the history of Massachusetts. There's John Adams, there's the Adams family, there's the Massachusetts Constitution, there's his relationship with Thomas Jefferson, all of these aspects of a single man, which then kind of filter out through the entire history of the state. Am I getting that right? Is that the kind of thing –

MAISLIN: Yes, yes. And in fact, it's more than just connection. It's editorial connection. It's the human ability to combine or uncombine things. The terms they use in library science are lumpers and splitters. Lumpers take ideas that are somewhat related and shove them together into the same package, such as, well, these are red things, even though they might be very disparate objects. They're all red and I'm going to lump them together under red. And then there's splitting, saying, well, there are all these different shades of red and there's an important distinction that needs to be made.

But it's editorial. You could take an example in a computer book where the concept of windows actually means a couple of different things, Windows the operating system, which is built and sold by Microsoft. You have windows which could refer to the user interface. And in a computer book, both of those words tend to show up. Instead of lumping them together to say, well, they're connected

because they're spelled the same, you actually have to consider that they're not connected because they mean different things. A computer-generated index can't do that. It sees the word windows and it knows that the context is computers, and it still can't make the distinction that a human can make.

KENNEALLY: It's kind of like spell check. You just can't rely on it at all.

MAISLIN: That's right. Or the argument which I've heard is you can rely on it, say, 80%, or if you have a really good spell checker, 90%. The problem is you can't find the 10% that's wrong unless you go through it with the same attention that you would have given it had you just done it from scratch.

KENNEALLY: Now, splitting, is that the same thing as creating subentries? Why don't you talk about that a little bit.

MAISLIN: Sure. It's not the same, although subentries are a tool that you can use for splitting. Splitting, if you think of it, is just the theoretical concept of saying, these are different things. I need to show that they're different things. But the way I look at subentries, it's sort of counterintuitive. In fact, it's counterintuitive to almost every indexing textbook that's out there. They'll talk about when to create subentries and so forth, and my argument is that subentries, as cool as they are – if you've never written an index before, there are two things you're all excited to do. One is alphabetize and the other is make subentries, because if anything is an index, it's an alphabetized subentry thing. But the reality is subentries is just a convention for making things more visible. You could take every subentry in an index and run it up to the main entry. So instead of having a main entry, for example, computers, and a subentry for configuring, you could easily create a single entry that's 'computers, configuring' or 'computer configuration.'

In a way, subentries don't provide any additional access just because they're subentries, but the advantage of subentries is it allows you – in a sense, maybe it's more of a lumping tool. You can create a main entry for computers and you know that all the subentries are related to that large topic. So in fact, maybe it's the opposite of what you're suggesting, that it doesn't split, but it's a way of grouping things, in a way, and in that sense, it's more of a lumping tool.

Subentries are wonderfully visual. They're often a big feature of indexes. But the reality is that they're overrated. Knowing how to create subentries and how to write them is extremely important, and an index that doesn't have subentries is probably flawed. But the reality is that the subentry nature of it is really a formatting convention that doesn't have nearly as much value as the language that you choose, whether or not to include something in the index, how you categorize things and so on. In that sense, I do. I think a lot of energy is placed on subentries where I think it could be better placed on some of the other theories and practices of indexing.

KENNEALLY: That seems like a good and practical tip, and we're about to get to the point where we want to talk about doing your own book, whether you're an author or a publisher. You did have a step one that you told me. It sounds like it's a pretty easy place to begin.

MAISLIN: Sure. I think if you had to describe indexing and what I like to think of as my three easy steps, is that first, you have figure out what goes in the index. Then you have to figure out what to call those things, what words to use. And then you have to figure out how to organize them. The organizing is the subentry piece. You can see it's the third step. It's something that can only be done after you've done the other first two.

The first step and the place where you can really separate the people who know what they're doing from the people who don't know what they're doing is what you include in the index. And I have to say that authors tend to fall into the camp of not knowing what they're doing. It's unfortunate, but it's because of the bias that authors bring in, and I'll explain that in just a minute.

The first step of indexing is knowing what goes in the index. So as an indexer, you have the pages in front of you. Or even if you're building an index of web pages, you have to have the web page in front of you to decide what kind of keywords you want to use to make it show up in Google or your own company search engine or whatever. And if you put every word in, it's a little bit more of a concordance. So you're looking for the things that people want to look up.

During this call, for example, I could make a joke, and I would consider the joke not something you'd want to look up, even if anecdotally it's somehow important. People aren't going to go back and look up that joke, so you don't index it. In that sense, you don't burden the reader with entries that point to something they don't want. And I suspect anyone who's ever used an index has had the experience of going to the index, looking something up, turning to that page and going, why am I even here? This annoys me. And in fact, that's one of the criteria you use. You don't index something that if found, will annoy the person who looked it up.

So you have to be particular, which means you have to know your audience, you have to know your subject matter, and you have to have enough time to write the index. If you're rushing, which is often the case, you don't have the opportunity to really extract the important things.

You could look at, for example, even just if you were indexing television shows, the odds are you're going to index the show as a whole and you're not going to index the quotes or the character names or the plot lines. You're going to index the show as a whole.

So knowing how deep to go extracting is a big challenge, and people who are just starting in indexing or haven't taken any training or really thought about indexing,

this is the hardest thing to learn. It's that not indexing is the more valuable skill. You don't include things. You have to be OK with skipping an entire paragraph and not indexing it, because nobody wants to look at that paragraph out of context.

KENNEALLY: What would be –

MAISLIN: An example of this would be, let's say you have a procedure. It's a 12-step procedure. Step 1, log in. Step 2, type in your password. Those kinds of things. The individual steps don't have any valuable information. If I look up log in in the index and I turn to a step that says, log in to your computer, that would annoy me. But I'm not saying that the steps themselves aren't important. If an author wrote a 12-step process in only nine steps, something would be wrong. So indexing is about indexing what's important, and this is where authors have that challenge is that most authors see what they write as important because – and I don't mean this is in biased or an egotistical way, but in a very practical sense. You don't write something if it doesn't need to be written. So there's the sense that, well, if I wrote it for a reason, I'd better index it. And the fact is, it's not true. Indexing is something different.

KENNEALLY: I was thinking, though, of a quote by Mark Twain that goes something like, what's important about writing is not what you write, but what you leave out, and I think he was referring to the need that writers have to edit their own work and kind of pare things down and have the book reveal itself to them in the writing and the rewriting that goes on. But as you say, at some point, when what's left is the diamond out of that lump of coal, for the author, the creator, that's all valuable, and you really are trying to look at it from the reader's perspective, the consumer. I hate that phrase. We talk about content consumers, information consumers, but that's really what it comes down to.

MAISLIN: You're absolutely right. If you ask people why they use the search engine that they do, and let's suppose it's Google, which right now is the most popular engine. Why do you use Google? And people will say, because it thinks like me. Which means they believe already that there's some editorial process taking place. But then if you say, well, no, I mean like practically. How do you know it thinks like you? And they say, oh, because what I want is usually in the first couple of results. Which means what they're interested in is the first couple of results. The fact that Google can turn around and give you 450,000 pages is irrelevant. That's a lot of garbage for most people. In fact, Google doesn't even allow you to browse all 450,000 if you wanted to. It only allows you I think the first thousand or so. People don't want that. People like Google because they like the first page of results. That's what they're after. Indexing is the same way. It's the editorial process of ignoring 449,000 choices to make sure that what remains is what people really want.

KENNEALLY: And so how do you then – what are some tools, what are some techniques to learn what to ignore and what to pay attention to? You've told me

about some of them, that it's as simple as the page numbers, but then there's various references and language choices. Talk about some of those pieces.

MAISLIN: Well, it's a challenge. Knowing what to index and what not to index is a lot of gut feeling that's based on experience, and people who have a knack for indexing just kind of get it. If you've ever had a conversation with somebody who just doesn't know when to stop talking, that's not the kind of person you want indexing because they don't know when to let go. So to some extent, a lot of good indexers might be very meticulous and practical at the same time, practical in the sense where if they spend too much time working on it, they're probably over-indexing.

But there is no good guideline except some kind of usability test, and the key is knowing your audience. Will somebody look this up? That's the first piece. But once you've identified what people might look up, you go to that second step of indexing, which is what words do you use? What do you call it? And this is where you could take a single item that you think is that diamond and find a way to include it in the index eight, nine, ten different ways with different words.

KENNEALLY: Now, is that like in your example with the bread in the Bible?

MAISLIN: Yes.

KENNEALLY: Synonyms and various kinds of metaphorical meaning and all of that?

MAISLIN: Yes. Fundamentally, there are what I like to think of as four different categories of ways in which you can index a single concept. You can index it by what it's called. In this case, bread. You can index it by synonyms, and I can't think of any synonyms for bread right now.

KENNEALLY: There's synonyms. Well, you're right. But maybe like toast and crackers and –

MAISLIN: Right. Well, then there are variations on the word and there are subsets of the word and parents of the word, so grain products might be a parent category and toast might be a subcategory and sandwiches might be an application of bread, so that's one area.

And then of course, in the case of the Bible, you also have categories that are symbolic as well as actual denotative things, like toast and bread are clearly related, but bread and nourishment is not an obvious relationship. But if you know the audience and you know the subject matter, you can get that out of the book.

And then on top of that, I like to think of the fourth category as all the things that are wrong, but it's what people actually do. So you could argue that if I'm talking about bread and somebody looks up toast – no, I'm not talking about toast. I'm

talking about bread. Or somebody looks up baguette. No, that's too specific. I just look up bread. It's just bread. So in a way, you have to be thinking of what people are looking up. You have to be aware of their language as well.

One of the challenges of writing a book for the layperson is you don't know what their words are, and you have to put yourself in their shoes, especially for technical books or scholarly books where they don't even have the language, or maybe even they don't know how to think about the subject and they're going to have to look it up anyway and you have to provide enough access at that point.

Thankfully, you're only doing this for the diamonds. If you are over-indexing, if you're indexing all the chaff and then on top of that, trying to put a whole bunch of words to it, your index would be mostly unusable. So you find those gems that you really want to index and then you index them to death based on who your readership is. And that gives you essentially the word building blocks of your index, and that leads to the third step of how you organize it, which comes back to the subentries that we were talking about.

KENNEALLY: And something like – you mentioned baguette, and I was thinking all the types of bread, baguette and sourdough and challah and whatever else there is. Is that an instance where you might want to just say, 'baguette, see bread'?

MAISLIN: Yes. And you can also go the other way around. You can say 'bread, see also specific kinds of bread.' So instead of having to list them all, you could create other tools.

Cross references, like subentries, are an access to other pointers. You start in the index some letter of the alphabet, B for bread, C for challah, H for people who don't know how to spell challah, and then from there, you provide page numbers or subentries or cross references or some combination of all of those things to ultimately point them where they want to go. Cross references are a tool that allows for that kind of thing.

Page numbers and cross references do different things, but ultimately, they have the same kind of functionality, which is, I'm going to get you that much closer to what it is that you're looking up.

KENNEALLY: We are talking right now with Seth Maislin about the basics of good indexing. My name is Chris Kenneally, the host of Beyond the Book from Copyright Clearance Center. We're going to take a quick break and be right back for the second part of our conversation.

(pause)

KENNEALLY: And we are back right now with Seth Maislin, the immediate past president of the American Society of Indexers, talking about a subject that I think,

judging by my own sense of excitement in this first half, is something that surprises us all when we begin to think about it, because we know a lot about indexing, but we never really think about how to create one and what goes into all of it. And that's what we're doing here with Seth today on this special program for Beyond the Book. And really what we're trying to get at is how an index helps authors and publishers alike leverage the power of their information, and welcome back, Seth.

MAISLIN: Thank you.

KENNEALLY: We did a survey of the people participating in the call, and I can share with you some of those results, because I think they're revealing about how we do have our antenna up about indexes. We asked people how important they think an index is to published work, and you'll be gratified, Seth, to learn that 95% think it's an important element. But here's something that may disappoint you, at least professionally speaking. We asked them to rate indexes for the works that they've written or published and many of them, in fact 50% – 52% – said they found them fair, fairly done, not so great. Only 36% said they were good, and only 10% said they were excellent. So while everyone thinks an index is important, many people are not terribly satisfied with the indexes that they have in their works, whether they are authors or publishers.

So let's talk a little bit more about what makes a good index and a bad index. You've talked about what's not indexable and knowing when to stop and what's important, but how can I recognize, if I'm going to go hire someone to do this for me, whether they're going to do the job well or not?

MAISLIN: Those are good questions. If you have the ability – let's suppose you have three index portfolios. You're looking at indexes written by three different people who may or may not be professional indexers, but you're just looking at their work. Or similarly, you're in that bookstore scenario where you want to look at the index in the back of the book and you want to be able to just judge does this book have a good index or not? You're basically looking to see whether the analysis and discrimination and language choices all make sense for the book and for you as a reader.

One of the reasons I think there are so many indexes that get poor ratings is often because of the way the production process works. You might have somebody who is not a professional indexer writing it. Often indexers, even if they're really good, indexers are forced to write under ridiculous deadlines because that's the nature of the production process. The index is the last thing and suffers because of delays in the production process that took place beforehand. Or maybe space constraints. The cover is often printed before the pages are printed, and if the index is running too long to actually fit inside the binding, it might get cut back. Those are production reasons that can cause anything to suffer, not just the index.

But when it comes to the analysis of an index, the types of things that you see that go wrong are somebody who indexes something that really doesn't need to be there. And the only way to determine that is to randomly spot check an index. I'm going to look something up. I'm going to turn to that page, and if I think the indexer knew what they were doing, I'll know right there. Do I go to this page and find what I'm expecting to find?

Some other issues that tend to come up that might be more obvious have to do with sort of the ratio of what I like to think of different page numbers to different subentries. So people who love subentries – and don't we all? – people who create lots and lots of extra subentries might very well take something that's on one page and give it three or four subentries even though it's all on the same page. This is the kind of mistake you see based on inexperience. So imagine an entry that has four subentries and imagine that all of those subentries point to the same page. That's an extreme case, but clearly what it's doing is it's putting a burden on the reader. The reader somehow has to – you're being asked to choose among these four subentries, and then no matter what you choose, you get the same result.

So when I think of the ratio, that's what I'm talking about. If I have a book that I'm pointing to a topic that runs across 20 pages, but I don't provide any subentries, well, that doesn't help me get into the 20 pages. But if I'm looking at something that runs two pages and I'm giving you eight ways in, that's kind of overdoing it. So that's the ratio that you're looking at, and that's the ability of somebody to use subentries.

There are also language issues. So does the index provide me the language that I use in my head? If it's author-written, often authors are working very hard to constrain their language. They want to use a word like algorithm. It has a very specific meaning for them. That's the word they're going to use. It's not a process. It's a tool. It's not a formula. It's an algorithm, and so they work really hard to use that word and that word only in their writing. When they write the index, they've already kind of filtered out all that other language. I may not have the word algorithm in my head. I might think of it as a formula or an equation, and if those words aren't in the index, the index has failed me.

So the ability to think outside the book is one of the tools of an index. Does it have the words that you use? And then when the words are in there, do they make sense? Do the main entries and the subentries flow together to create an easy idea that you can get in your head? And are the words in there actually accurate to what it is that you're pointing to?

In fact, one of the errors that's unfortunately quite common is actually just the page number itself is wrong. It tells you to turn to page 56, but guess what, it's actually on page 58, and you have no idea when you read it, and that's an error. That can be a production error, but these are all the different kinds of things that can damage the quality of an index and they're not uncommon.

KENNEALLY: It strikes me, Seth, and this is something that came up in some of the questions that people submitted to us before the program began. We live in a world where we think that technology, the computer, generally speaking, can solve all kinds of problems and we don't need human beings to take care of this job, we can get the computer to do it. Computers are replacing car mechanics. They're replacing everybody. Is there a concern that there are software tools that are out there that are sufficient replacements for indexers and we don't really need a human being and – I'm not asking you to kind of defend the profession here, but it strikes me that that's an important thing to bring up here, that the human element, the choices, the experience, the background knowledge, these are all things that cannot be part of a piece of software.

MAISLIN: I think the challenge is that a computer is only as good as the information you give it, and often, computers that are attempting to index are only looking at some very basic things like the word and its spelling.

Can you look at a book and – how well do you know that what's in that – when you read a book and you know a little bit about grammar, it's pretty obvious to recognize when a word is a noun. Computers don't do that very well. And in fact if they're only looking at the words and not the linguistics, then they can't tell the difference at all. Now, you could build linguistics into a computer and you could say, oh, well, now it's a smarter machine, and you could put a thesaurus in there and say, well, now it's a smarter machine. And in theory, you could build a computer that indexes that is so obviously good that you can't tell the difference between the computer and the person.

But think about the energy that's required to build that tool, and this is why it's theoretical and quite implausible. And that's that if you're trying to create a computer that thinks like a human, you have to program it with all the subtleties and the context and the knowledge that you have as a human, and the energy that you put into building that tool is much better spent just writing the index yourself.

So there's a limit to what computers can do. I often think that even if you sent a computer into a book and you say, I just want to extract the names. If there's a name in this book, I want a computer to find it. What tool do you use? If you look for things that are misspelled because, well, it's a name, so it's misspelled, but everything else is correctly spelled, that's great until you get a name like Prudence or John, names that, while they're correctly spelled English words as well, they mean something else. And then you say, OK, well then maybe it has to be capitalized. Well, the fact is is the Central Intelligence Agency is a series of capitalized words. How do you know that's not a person's name? How do you know the name of a town? Baton Rouge? Well, Baton's spelled correctly. Rouge isn't, so maybe it's a name, but nope, it's not a person. And then there are names that aren't capitalized.

But where it starts to get even uglier is the fact that one person can have lots and lots of different names. You can have President Clinton and you can have President William Jefferson Clinton and William Clinton and Bill Clinton, and all those happen to be the same. So even if the computer could extract all the names, it doesn't know when it's the same person or not.

KENNEALLY: And even in those examples, that's relatively easy, but I was thinking about – I don't know why we both thought of presidents, but there was President Eisenhower who was also General Eisenhower who was also Ike.

MAISLIN: That's right. And then you have parents and children who have the same names, and then there's a coincidence. I worked at a company where there was a Ronald Reagan during the Reagan administration, so you had two Ronald Reagans who were different people, and there's no way to know that, and so on. And even now, you talk about the Bush administration and you need to know the context. Which Bush are you talking about?

So even if you could get a computer to do all of that, the only way to really do it is to still say, well, it's 80% correct. I'll go back and look over the other 20%. And it's easy to look through a list of names and say, Baton Rouge, no, that's not a name. But it's hard to go back through the book and see what you missed.

And the one thing that's missing in this example is the very first step that I gave earlier, which is, is it important enough to look up? There's no analysis. Just because I talk about the Bush administration doesn't mean I have anything to say about the Bush administration. It's very to say, in the early years of the Bush administration, something happened with computers. Well, that has nothing to do with the Bush administration, yet, oh, there's a name, which by the way is an actual English name, Bush. It's a word in our language. It's capitalized, but that doesn't say anything. So how are you supposed to find that and then on top of that know not to index it? And that's where the human being – we just have this ability. It's an innate, cognitive ability to know what's important and what isn't.

You can step into a room and you know immediately with a quick scan if there are any people in the room, but believe it or not, computers can't do that because they don't know the difference between what's close and what's far without specialized equipment. They don't know that it's a person, and if the person moves, maybe that's a hint, but if the wind is blowing, even paper moves. Something that simple we just have the ability to do, and that's why people will always write better indexes than computers.

The question is, are you OK with 80% and does that save you enough money that it's worth it? And when you're talking about books and even encyclopedia – you're talking about printed, readable things – no, it doesn't make sense. Only in an environment like Google where you're searching billions of things – it just doesn't make sense to say, oh, I'm going to hire a million indexers to work on this

around the clock. OK, you'll use the computer and it will be good enough, and yes, you'll get 450,000 hits, but for the most part, what you want is up at the top. Of course, then you don't know if it's relevant or accurate, but hey, that's good enough because I have no other way to search billions of documents. That's where computers pay off.

KENNEALLY: And Google even addresses that because it said, OK, we can search for these words, but the way we're going to put some value to them is how often they're linked to other pages, so we're going to assume that that's a value. They had to have some way to sort it and they went that way.

MAISLIN: Yes. Now, I should, just to clarify – we're talking – there are two different classes of indexing tools. There's the automatic indexer, which is a misnomer. It's more of a concordance-builder. Even if you smarten it up with all sorts of knowledge, it's a concordance-builder, so the machine that writes the index. And those are almost 100% predominantly bad, and I say almost because there are those in environments that tend to function very well in giant document scopes like Google, which does a pretty decent job for what it's trying to do. It's not perfect. It has a lot of flaws that may not be obvious to most people, but there you are.

Then there are tools that enable a human to do his or her job better, and I'm talking about – there are indexing software tools that are specific to what indexing is. This is like what a word processor is to a writer, this is what the tools are for somebody like me, and there are a few of those that exist.

And then on top of that, there are documentation programs like Microsoft Word or Adobe Frame Maker that have some kind of indexing functionality in it. The fact is is that those tools, like Frame Maker and Word and WordPerfect and Quark and other things that have some indexing functionality, it's unfortunate, but the majority of those tools – in fact, I'm going to say all of those tools – are inferior when it comes to indexing. And the reason for that is that the tool isn't designed as an indexer. It's designed as something else. It's designed as a word processor. It's designed as a document management system. It's designed as a layout program.

KENNEALLY: So it's an afterthought to have that.

MAISLIN: It's not just an afterthought, but people aren't going to say, oh, I'm not going to buy this because the indexing stinks. So in a way, there's not even a consumer demand to have these things improved, so I think they will always be substandard. I don't think there's a market motivation for these programs to be improved by the companies that make them. And I think if you are going to use that to index – which might make sense in an overall system. If you're in document management, it makes sense to index using that tool – you'll always have to find a way around the missing features that you would get in a dedicated program designed for indexers for indexing.

KENNEALLY: That warning is probably a good place to stop our own little conversation here and include the other human beings on the line. Operator, I wonder if we can ask people to indicate that they have some questions. Again, the way to do that, to get in line, if you will, is to press star, one on your telephone keypad, and that should bring the operator in. Operator Cory, do we have you with that?

OPERATOR: Absolutely. If you'd like to ask a question at this time, please press star, one on your touchtone phone. To withdraw your request, you may press star, two. Again, to ask a question, please press star, one at this time.

KENNEALLY: Seth, while we're waiting for a question to come in that way, let me ask you about something that someone did write in to us on the survey and you alluded to briefly in your own experience working with Mercedes Benz, and that is that in the German publishing culture, indexes are not very common at all. This person was asking us about translating indexes and what techniques could improve the ability of an index to be translated. Can you address that?

MAISLIN: Yes. I can say from my perspective that unequivocally you should not translate an index. You wouldn't translate an index because what makes an index functional when I talk about context, includes things like cultural context. The way we put words together, the way we relate concepts, it's a very cultural thing. And maybe, again, for 80% of what you're looking at, that may not seem true, especially if you're looking at languages that come from the same basic roots or cultures that are fundamentally the same. But there's a big challenge with translating because in an index, you don't see the context of what you're writing. But again, there's a cultural context that's being there.

Now, I say that as saying, never, never translate an index. Always find somebody to write an index from scratch in the language that the book itself is written or translated into. It's easier to translate a paragraph because you know what it's saying than it is an index, which is just a bunch of words.

That said, if you talk to a translation company, they will tell you, no, we're really good at it. We know enough about indexing and a lot about translation and we can do a good job. So I will admit there are differences of opinions, but I have to say that as an indexer, it's a bad idea, which is not to say that translators can't do a decent job of it, given the opportunity.

I really think that if the goal is to write a quality index, you don't want to go down that road. If the goal is to at least have something in the back of the book, OK, sure. I get it for practical reasons, but I generally think translation just is not a good approach.

KENNEALLY: OK. Good. Now, we have an e-mail question here and it's asking the question, how do I become an indexer and what is the outlook for indexing as a profession? I know we have a lot of people on the call who are writers, authors of

one type or another, and they may think they've got some skills here that they could apply in the indexing profession. What's the outlook and how difficult is it to break in?

MAISLIN: I think that the outlook is good. Fundamentally, it is good. We are still in an information age. I suppose people will tell you, no, we're in the nanotechnology age, but I think we're still in the information age. All that talk about paperless society, we still actually print on more paper than we ever used to. And as information is now being created by everybody in terms of blogs and wikis and everything, there's just such a flood of information – an information overload – that the need to organize it in one way or another is just growing. So in that sense, the need for indexing skills is absolutely 100% increasing.

Now, if you want to index just books, there's still a solid market for that. There are still more books published now than there ever were. Even though some of them are going into CD-ROM formats or some people are moving online, it still needs to have an index. So it's still there.

There are a number of opportunities to learn indexing. The one that I just know – as you mentioned, I'm past president of the American Society of Indexers – is that there's a course that the American Society of Indexers offers. It's broken into three parts. It's reasonably priced. And it's a good way, starting from the beginning, to see, do I really want to do this? You can take the first part and see if – do you think like an indexer, and if you're interested, you go down that route. And there are other courses as well, but that's the one that I'm most familiar with, with of course the exception of my own, which I'm not here to sell. And that's also the most accessible.

The market is out there. If you're going to be a freelancer, it can often feel like a challenge to break in. I would have to say that on average, if you're serious about it and you're prepared to do the marketing that you need to do to find clients and get out there, you could get a business going in under a year. And that's on average. Some people, the first day and they already have connections and they're moving in.

And the amount of money it takes to start being an indexer is negligible. You take a course, you buy the software and everything else, free business cards, free e-mail. You already have a computer at home. You can get the business started and running under \$2000 U.S., and your first project might very well be worth more than \$2000.

So it isn't hard to get started. It just requires a certain amount of discipline and I strongly recommend some form of training. Not, I'm going to buy a book and read about it. I mean some actual course material with some kind of instructor.

KENNEALLY: Just again, Seth referred to the American Society of Indexers, and their website is asindexing.org, and you should go there if you're interested in the profession to learn about the courses they have and just background on the job outlook there.

We have about five minutes left on the call. I want to encourage anybody online with a question for Seth Maislin about indexing to press star, one on their telephone keypad and that will put you in touch with the operator.

Operator, do we have anybody online with a question? We have a rapt audience, then.

MAISLIN: I've said everything.

KENNEALLY: I don't know about that. I think, frankly, we've only just gotten started. It may be that it's a bit overwhelming for people and maybe they're just enjoying listening.

Let me ask you something that you mentioned about the growth of information. There's a certain interest in all of that out there with what is known as search engine optimization, which is a kind of indexing. Talk about that as a special skill. How is it different from the book indexing?

MAISLIN: The fundamental difference between what is called search engine optimization and indexing – and I should point out that search engine optimization is a bit of a misnomer. The goal of search engine optimization is to have something show up at the top of Google, so the goals there are market-based and not about finding things that you might want so much as about pushing things that you want people to find. So it's a bit of a flip.

So if you think about that from a book indexing perspective, if there's something that I personally as the indexer think, wow, I want to make sure everybody finds this, then the best way to do that is index nothing except that paragraph and then index that paragraph eight, nine, 12 different ways. And if you think about that as an index, well, that stinks. That's not an index. That's terrible. But search engine optimization is exactly that. It's a marketing tool. It's spam, in some ways, even if it's targeted spam. What is the best way to make something that I want found found? It means you're not considering the audience at all in that decision.

KENNEALLY: In fact, it's the opposite, it sounds like, because you were talking there at the beginning of our program with what you leave out, and search engine optimization, it wants to put as much out as possible.

MAISLIN: Yes. In fact, you're being paid to put something out there, and there is an actual book indexing example of this. There are university presses that will demand that the indexer include every name that's in the book in the index, even if

it's not relevant, like I gave you the Bush administration example earlier. And the reason they do that is they know that if a scholarly person sees their name in the index, they're likely to buy the book. It means there's a financial reason to include things in the index that really aren't relevant to most readers. They're just relevant to that person. It's like how many people on this call are part of my family? They know nothing about indexing but they're all excited because Seth is on the phone. And in a way, if I were to run my life that way, it would seem pretty backwards. I wouldn't be appealing to a wide audience. I'd be appealing to the people I already know. So indexing is a different approach, but a lot of the skills do overlap between the search engine optimization and indexing.

There is one other topic I wanted make sure that I threw out for myself, and that is the idea where if you don't want to index your book and you want to find an indexer, there are a number of ways to do that. Again, you could go to the American Society of Indexers, and you could even perform a search at Google and type in indexer or indexing and some subject matters. But there are a lot of indexers who are out there. We have more work than we know what to do with, which is a good thing for marketing, but it also means that we generally have a lot of experience. It's a small industry but a very specialized one.

KENNEALLY: We have an e-mail question that kind of addresses some of that, and you have already talked about the potential for finding work. Someone here is asking us, how do you estimate a project when you're contacted by Mercedes Benz or whoever it is? How would you go about sizing things up and knowing what this is eventually going to take as far as hours and then therefore how much to charge?

MAISLIN: That's a beautiful question, and I will say that there's a certain art that goes to this. The indexer, to some extent, is looking to make decent money for the work. The publisher or the author who's contracting is looking for a way to lower the price but still get a quality job, so there tends to be a negotiation that always takes place.

But if you're looking for ballpark numbers, the kinds of things that you have to pay attention to are how dense is the book from an indexability point of view. That is, if I look at a page, how many things on that page might somebody want to look up? So a college textbook for example, especially a science textbook, might have a lot of things on the page, so the price is going to go up. Whereas something that's very narrative in its reading might have less to index because you really want to index passages and not small ideas.

The deadline is a big piece. If you gave me a book and you didn't give me enough time to do the job, I have to charge a lot of money because in know I'm going to have to be working nights and weekends just to get the job done.

KENNEALLY: There's a rush charge there.

MAISLIN: Exactly. There are rush charges involved. If you're requiring specialized tools, that can raise the price.

But fundamentally, if you're looking for just average numbers – and I don't know what you really consider an average book because everybody has a different impression. But generally speaking, most indexers charge by the page, and I'm talking the book page. So if I have a 500-page book, I'm going to get 500 times my rate. Again, deadlines, density, how difficult the material is, what kind of audience it's for, tools, all of those things affect the numbers, but sort of the median numbers in there tend to run from about \$2.50 a page on the low end to as high as \$5 or \$6 a page on the high end, and even then, there's the possibility to go lower and higher for some books. I've seen coffee table books for airplanes, which are mostly pictures, \$2.50 is probably a lot. On the other hand, I've seen rush jobs for Microsoft that are extremely dense computer language for high-end administrators, and that's the kind of thing where it might be a very high profile book as well, and those can go much higher than \$6 a page.

The indexer tends to make money when it comes to speed, in that case. Do you know what you're doing? Do you know the tool? Are you specialized? Do you have experience? And that's where you can turn, as an indexer, that per-page rate, into a very good hourly rate without, in a sense, hurting the person on the other end who's figured out what the whole cost of the index is going to be.

So that's kind of my algorithm. If it's a 300-page book that's kind of an average textbook, it's probably about \$3.50, \$4.00, \$4.50 a page, but again, what's an average textbook? And if an indexer has done enough work, they can do a better job of recognizing, oh, that index is going to be more than it should be.

KENNEALLY: Well, listen. It has been a pleasure for me to speak today with Seth Maislin, who is the immediate past president of the American Society of Indexers and is a managing partner at the indexing firm, Potomac Indexing. Seth, thank you very much indeed for joining us today on Beyond the Book.

MAISLIN: It is my pleasure.

KENNEALLY: If you would like to review any of the material that we have presented during this call, the best thing for you to do is to go visit www.beyondthebook.com. Later this fall, we will have the program available as a podcast. You can listen at your convenience again. You don't need an iPod or an MP3 player. Most of the programs – in fact, all of the programs – can be played directly at your PC or downloaded and burned to a standard audio compact disk.

We also provide transcripts to make it easier for you to find, and perhaps, Seth, we'll get you to do the index for the transcript there.

And finally, we encourage you to subscribe to the entire series of programs we do. We cover the whole range of topics involved in the business of writing and publishing, from marketing to research and just about everything now, including indexing. Again, that address is beyondthebook.com.

If you have any questions, comments, or suggestions for future programming, please contact me directly. My name is Chris Kenneally and you can write to me at Copyright Clearance Center, [chrisk – C-H-R-I-S-K @copyright.com](mailto:chrisk - C-H-R-I-S-K @copyright.com).

Thank you all for joining us and, we hope you'll be with us again on a future program for Beyond the Book. Goodbye now.

END OF FILE