

### Introduction

Today it is entirely possible to create an index solely using digital devices but this is a very recent development in the centuries-long history of this particular information retrieval tool. This article discusses the mechanisms of creating indexes from ancient to modern times. The focus is on stand-alone books rather than journals, databases or websites but many of the same principles apply to those also.

### Why create an index?

Indexes are tools to enable the content of a work to be accessed in a way other than reading it in entirety; a means to swoop in at a precise point to find certain information. People have always wanted to be able to find, or re-find, a particular topic within a longer piece. Wellisch (1983) says that:

*the root of the word [index] expresses a fundamental communicative action that transcends all languages and probably preceded them in the evolution of man: to show something to another human being by pointing to it with an outstretched forefinger.*

### Who creates an index?

In the very earliest instances of indexes it was often the owner, or at least reader, of the book who created the index, in much the same way that we often create our own mini-indexes to recipe books or family photo albums on our own shelves today. These days it is usually the author, a professional indexer or someone at the publishing house/academic department who is tasked with the job.

The indexer is a shadowy sort, rarely acknowledged and it is not at all uncommon for a professional indexer, on confessing their chosen career, to be met with a delighted 'Oh, I never realised someone actually did that'. There are few famous indexers and only one known martyr to the cause, William Prynne who had his ears cropped for his all-too-revelatory index to *Histrio-Matrix* in 1634 (Archibald 2014). G. Norman Knight (1968) documents the emergence of the professional indexer in the 18<sup>th</sup> century, describing their '*inferior status – a Grub St Hack*'.

Professional indexers these days, however, are highly trained, often freelance and use a variety of tools, both specialist and generalist, to create their products. Today there are directories of professional indexers and Hazel Bell's book of mini-biographies is enlightening if you want to find out more about the kind of person who takes it up as a job (Bell 2008).

### Underlying analytical skills

A good index points to all the useful information within the text, without being bloated by references to irrelevancies. Therefore, no matter who creates an index, the basic underlying analytical skills are the same regardless of the technicalities of producing it.

The skill of the indexer is in the small, strategic decisions which cumulatively make an effective index.

Archibald (2014) says:

*to do it well requires an ability to quickly synthesize information and conceptualize the thematic relationships within that information [...] A good indexer excels at communicative terseness [and] the golden mean is an indexer who is objective enough to anticipate the needs of a wide variety of readers and subjective enough to parse the central importance of the work*

So first the indexer must identify names and concepts, usually by reading the work cover-to-cover and always with the reader in mind. Secondly they choose a word or words to describe them, which may or may not be identical to the phrasing in the text. Thirdly the indexer must find a way of indicating the destination point; a 'locator'.

### Physical creation of the index

Beare (2007) looks back to the very first pictorial images in cave art and hieroglyphics which include elements of indexing, likewise do the 3,000-year-old *I Ching* (Humphreys 2011) and ancient Hebrew texts (Weinberg 1999). So we know that for probably thousands of years people have considered the labour that goes into producing some form of an index a worthwhile investment in the work's future. Handwritten scrolls, some of the earliest known forms of long texts, did not lend themselves easily to indexing but the codex, which is more book-like in nature, certainly did.

There is debate over when the first subject (rather than name) indexes came to be produced. Ancient Roman Seneca provided Lucilius with a page of notes directing him to certain sections so that he wouldn't have to read the entire work, for example (Archibald 2014). Subject indexes also by definition trace back to the origins of alphabetisation itself (otherwise many early 'indexes' are rather better described in modern terms as 'tables of contents'). This goes back to at least Alexandria in 3rd century BC Egypt when Callimachus created a catalogue for the famous library which was divided into subjects by alphabet (Wellisch 1994). There are alphabetical subject indexes to a collection of quotes and aphorisms called the *Apophthegmata* dating from the 6<sup>th</sup> century (Boardley 2018) and we know that there were many indexes to manuscript books. What we know very little about is the nitty-gritty of *how* these indexes were created, although we can imagine that the process was laborious and mentally taxing. Cornog (1983) notes that there is evidence that such indexes were often handwritten at speed, under time pressures that would be familiar today!

The growth of printing led to a real explosion in the prevalence of indexes although we still know little about the actual compilation practices in early days apart from that it was hard work. Wellisch (1994) talks about one of the very earliest printed indexes, the elaborate and complex *De arte praedicandi* of St Augustine, and says:

*the anonymous editor and indexer says in his preface that he had collated the text of several manuscripts which he had found in different places, though he may have improved on his models because he also claims that it had taken him a long time to compile the index.*

Duncan (forthcoming) quotes a fascinating exchange between John Oldmixon, creator of an index, and his client saying that it '*has cost me a great deal of Pains & richly deserves 121 [coins]*'.

In other words, good indexing takes time and effort, but it is worth it. *De arte praedicandi* is a book on the art of preaching and such a practical book would be relatively useless without a means to look up particular readings or themes. And if it is useless, who will buy it, or recommend it to their fellow preachers?

## Pages, paragraphs and folios

The question of what locators to use is interesting. It was only towards the end of the 15<sup>th</sup> century that foliation and pagination became common. Before that indexers used chapters, sections, paragraph numbers, marginal letters and an array of other means to indicate location (Wellisch 1986). In the earliest handwritten indexes the index-creator often had to write in their own page/folio/paragraph numbers – this had the advantage that, just as in the modern digital environment, the index could be copied when the document was reproduced even if the pages were different sizes. Line numbering was another common device.

## Card-based indexing practices

Throughout the 17<sup>th</sup> to 20<sup>th</sup> centuries indexes became a key part of scholarship and numbered pages became the *de facto* standard. In the pre-computer age indexers would have read the book, just as we do today, making paper notes as they went along as to concepts and names that needed referencing along with their locators. A common method was to use ‘index cards’, typically 5x4 inch cards with one concept/name per card which could then be reshuffled into an alphabetical order and the index copied up from the cards. Alternative methods did exist, such as thumb-tabbed notebooks (Stallybrass 1968). Jobbing freelancers were booked by letter or phone and hefty packs of paper proofs sent out with the resulting index mailed back along with a paper invoice and the hope of banknotes or a cheque by return. Wheatley (1902) describes the process of creating an index without a computer – cutting up slips of paper, mixing glue, alphabetising them by physically sorting them and hoping that a breeze doesn’t come in and disturb the arranged-but-as-yet-unstuck papers before copying the resulting index by hand or typewriter. Bell (2008) quotes a Margaret Anderson who said that ‘*the main enemies [of the indexer] are hurricanes, housegirls and cocktail parties*’.

Gradually with the advent of computers in the 20<sup>th</sup> century more and more of the physical labour moved to being handled by a computer and today’s enemies are more like broadband freezes, BACS errors and hard-drive failure.<sup>1</sup> Word processors came along first, enabling on-screen editing, spell-checking and even rudimentary A–Z ordering. Specialist indexing software produced a sea-change in how indexers work but also of importance for indexers were the advent of email, pdfs, Office-type software, broadband internet access and then digital publishing processes, tagging and embedding and online workflow processing.

## Specialist indexing software

The first specialist indexing software MACREX dates from 1981, a pre-Windows era (Beare 2004). The principles of entering content in a standard form and determining formatting, layout and sorting at the time of output (i.e. to a printer or a word-processing file) are common to all indexing software. There are three main indexing software providers, MACREX, CINDEK and SKY, and they share commonalities as well as having their own specialist features. All of them have developed iteratively in response to user requirements as well as the demands of the publishing industry and provide a variety of useful features such as sorting, formatting, error checking, spell checking and editing functionality.

Publishing software platforms such as InDesign, FrameMaker or LaTeX can have in-built indexing so the indexing is done directly on the live publisher’s file. Microsoft Word also has indexing capabilities although many indexers prefer not to work directly in Word itself but rather to use a specialist intermediary program such as DEXembed, WordEmbed or Index-Manager to provide either an embedded or tagged index.

---

1. Hard to deny that it sounds more fun in Anderson’s time.

Fully-automatic index programs are not yet available, or at least not ones that produce a usable index. However, semi-automatic programs using linguistic and statistical algorithms are becoming available, such as TExtract and PDF Index Generator. These create an initial draft from a pdf proof which can then be formatted, expanded and adjusted before reintegration into the text.

### Indexes embedded or tagged in the text

Up to this point we have looked principally at indexes created as separate documents which are bundled with the text at print stage. A subsequent development, however, is indexes which are embedded in the digital file itself or created using tagging procedures, including XML. This is used both for print books, allowing the index to be reproduced for e.g. hardback and paperback versions which have different page numbers, and for books designed to be used in a digital environment. In this case the index is click-through so there are questions to be debated about how to indicate where there is more than one mention, or the scale of mentions (e.g. the difference between a page range of 10 pages and a single reference). The index entries are linked to the text itself beneath the on-screen words in the form of hidden coding and can therefore be added earlier in the process than the traditional final page-proofs stage. The analytical skills remain the same though – the indexer must still identify what is salient, choose a word or phrase to describe it and then mark out a location to point to. Some publishers have developed their own proprietary tagging systems; others use commercial products.

### Conclusion

Indexes have been around for centuries for the simple reason that they're extremely useful. This brief overview has discussed how even though the basic analytical skills haven't changed, the tools and mechanics of index creation have transformed enormously with exciting new technologies emerging all the time. The future holds more developments in automated indexes, multilingual access, indexes across several publications and the continuing development of digital publishing.

### References

- Archibald, Sasha (2014) 'Indexes, in praise of.' *The Indexer* 32(4), December, 142–148.
- Beare, Geraldine (2004) 'MACREX – a history.' *The Indexer* 24(1), April, 18–20.
- Beare, Geraldine (2007) 'Past, present and future.' *The Indexer* 25(4), October, 257–264.
- Bell, Hazel (2008) *From Flock Beds to Professionalism: A History of Index-Makers*. Hertfordshire: Oak Knoll Press.
- Boardley, John (2018) *Index: A Brief History*. <https://ilovetypography.com/2018/08/24/a-brief-history-of-the-index/> accessed 28th August 2018.
- Cornog, Martha (1983) 'A history of indexing technology.' *The Indexer* 13(3), April, 152–157.
- Duncan, Dennis (forthcoming) *Table of Discontents: A History of the English Book Index*. Oxford: OUP (accessed via Society of Indexers' conference 2017).
- Humphreys, Nancy K. (2011) 'The world's oldest profession: indexing?' *The Indexer* 29(4), December, 161–165.
- Knight, G. Norman (1968) 'Book indexing in Great Britain.' *The Indexer* 6(1), Spring, 14–18.
- Stallybrass, Oliver (1968) 'An unusual method of making a book index.' *The Indexer* 6(1), Spring, 4–13.
- Weinberg, Bella Hass (1999) 'Indexes and religion: reflections on research in the history of indexes.' *The Indexer* 21(3), April, 111–119.
- Wellisch, Hans H. (1983) "'Index": the word, its history, meanings and usages.' *The Indexer* 13(3), April, 147–151.
- Wellisch, Hans H. (1986) 'The oldest printed indexes.' *The Indexer* 15(2), October, 73–82.
- Wellisch, Hans H. (1994) 'Incunabula indexes.' *The Indexer* 19(1), April, 3–12.
- Wheatley, H. (1902) *How to Make an Index*. London: Elliot Stock.