

indexing

1. Introduction

I would like to introduce you to our new Indexing Procedure.

We want to be able to take full advantage of the many different ways we can now market and sell Cambridge books, and we are constantly looking at ways to produce books more speedily and efficiently. At Cambridge University Press we have been developing a method of capturing our content electronically using XML (extensible mark-up language). The purpose of this is to enable us to capture fully the content of our books in XML so we can then repurpose this content (if need arises) into another form of product e.g. web, e-book, CD-rom, PDF etc. As a result I'm pleased to say we are now moving to a new procedure for the early stages of the production of the book (a new production workflow and with that a new method of producing your index), which will give much more flexibility to the content of the book and also speed up production without adding costs.

1.1 Here is a brief explanation of the production process that will lead to you producing your index.

- All components (text, illustrations etc.) of the book are now captured in XML at the very beginning of the process, before copy-editing. This basically means hidden electronic codes are put around the content.
- The output of this is an XML typescript which shows all components of the book, which looks like a slightly reformatted word file. The XML codes are not visible.
- The copy-editor will be sent a copy of the XML typescript to start copy-editing and you will be sent a duplicate XML typescript so that (a) you have a copy that matches what the copy-editor is working on to enable you to answer queries; (b) you can check all artwork (if any); and (c) you can produce your index.

1.2 What are the benefits of this new process, once copy-editing and indexing is complete?

- First page proofs will be of the entire book, including a fully set index in page form.
- There should be fewer corrections.
- We have a confirmed extent (length) for the book and are thus able to finalize the print run and price of the product sooner – a great benefit in marketing your book.
- The remaining stages proceeding to press should be faster and smoother.
- Overall schedule time should be shorter.
- At the end of the production process we will have a high-quality print product, and also fully XML-captured content for repurposing.
- We also have a fully usable index in whatever form the content is published in, now and in the future.

2. How does this indexing process work?

Most indexes need to refer to points and spans within a book's content. The 'points' are locations within the text which need to be referenced as having a single location – for a printed book they entries which refer the reader to a single phrase. The 'spans' are entries which refer the reader to a range of lines, paragraphs and pages. Index entries may refer readers to a mixture of many locations, both points and spans.

- Because XML does not use the page as its structural unit, there is no advantage in linking an entry to a page of the XML script.
- Instead, each index entry should be 'anchored' by marking the relevant word on the XML typescript, using a simple numbering system in the margin.
- Then, as you create your index typescript in the usual way, you will refer, for each entry, to the appropriate unique number.
- This will enable the typesetter to build a coding structure which will attach the word to its index entry and produce the correct page number for the entry automatically.
- The great advantage of this system is that, since each index entry or subentry will thus be anchored to its position in the electronic file rather than to its place on the page, this anchor will remain constant even if the book or electronic product is repaginated.

3. Creating your index in this new process

3.1 You will construct your index in exactly the same manner as you would a conventional page-linked index with the two exceptions, that you will not use page numbers, but a unique number as an anchor point for each index entry, and that you will use a highlighter pen or underlining on the text to indicate the word(s), or the start and finish of a span, referred to by the index entry. I enclose a sample chapter marked up and a corresponding index. You will see that on the first page of the chapter I have highlighted 'William Wyler's' as a single point and placed the number 2 next to it as the next available number. You will also see that I have indexed '*Wuthering Heights*' on the line above that as my first indexed item. I intend this to be a span and have placed next to it a unique number 1 but have added a 'b' (b for beginning) to indicate the opening of a span. You will see that further in the text, I have closed that span with a f1 (f for finish), and have highlighted the word at which the span is to finish. Both these numbers will automatically change to give the correct page numbers for both the single entry and the span once the text goes into page. If the text is altered, e.g. by the deletion of some material on the proof, the pagination and the corresponding numbers in the index will be automatically corrected to take account of this. If we wish to publish the text as an e-book in a completely different format from the printed book, again, the numbers in the index will be reconfigured automatically to match the 'pages' of the e-book.

3.3 You will see that the index entry '*Wuthering Heights*' has the number 7 as well as the number 1 next to it. That is because further into the text, *Wuthering Heights* is discussed again and requires a new entry in the index. The next and any subsequent unique numbers need to be added to the entry in exactly the same way as the page numbers are accumulated for a conventional index.

3.4 The unique anchor numbers do not have to be entered in numerical sequence. For example, on the first page of the chapter, the first epigraph source is Barthes, *S/Z*. I have already made an index entry for Barthes, and put in two anchor numbers, 3 and 15. I now want to add this new reference, which is number 24, and I can do so even though the anchor precedes the earliest number in the sequence – it does not matter that the sequence is apparently ‘out of order’ in this way. (Equally, I could make *S/Z* a subentry under ‘Barthes’ with the same number 24.) Having added this new reference to Barthes, I can also put in a reference to the second epigraph source, by giving it the unique anchor 25, and creating a new index entry for Cowper.

3.5 You can also, of course, put spans and point entries inside much larger spans. For example, the subject of the entire chapter is ‘eavesdropping’, and I have chosen to make a main entry ‘eavesdropping’, with number 26 as a span which encompasses most of the chapter. But I have also created subentries, both points (29) and spans (30) within that larger span.

3.6 It is also important to note that in using this system, there does not need to be a match between the index heading and a word in the text: it is the anchor number which creates the link between the two, not any sort of automatic word match. This applies especially to the indexing of concepts, and the entry on ‘eavesdropping’ is an example of this, as is ‘interpretation’, which I have used as the heading for the text phrase ‘interpretive activity’.

3.7 Some anchor points may refer to more than one index entry: see for example, ‘the Richardson-Burney tradition’: 22 is thus given as an entry for both Burney and Richardson.

Austen, Jane 10

First Impressions, see *Pride and Prejudice*
miscommunication in 14, 22
Persuasion 13
Pride and Prejudice 12, 16

Barthes, Roland 3, 15, 24

Brontë, Charlotte 9

Brontë, Emily 8, 1
Wuthering Heights 1, 7

Burney, Frances 19, 22
Cecilia 20
Evelina 21

Cowper, William 25

eavesdropping 26
as narrative device 28, 29
authenticity of information obtained 31
results in misunderstanding 27, 30

Farrer, Reginald 11

interpretation 6

judgement 17

McGregor, Graham 5

Miller, D.A. 4

Richardson, Samuel 22

Tanner, Tony 18

Wylar, William 2

Wuthering Heights, see Brontë, Emily

4. Your options

4.1 On hard copy or on screen.

You can either mark up a hard copy of the XML typescript and supply this with the index (supplied on disk with matching hard copy) or the Press can supply you with a PDF file of the XML typescript for you to mark-up on screen and supply the marked up PDF files with the index totally as electronic files.

If you choose the later option, you will need Adobe Acrobat 5 (the full program) and a good understanding of the 'tool' options in Reader 5 and how to use them. For those who would like to work electronically, I can provide you with detailed suggestions on how to approach this work.

Often hard copy is the fastest and easiest, but either way, we are pleased to be able to provide a hard copy or on screen option should it be required. Please let your production contact know which you would prefer.

4.2 Numbering throughout the XML typescript as one OR broken down into chapters.

You may find that the size of the index required takes the unique numbers into the high thousands, which you may find difficult to monitor. Any system of numbers may be used provided the numbers are unique. If you find that you are running into the high thousands you may find it helpful to start the series again with each new chapter: e.g. 1.27 = unique number 27 in chapter 1.

Another more sophisticated system which some of our indexers working electronically find helpful employs a combination of chapter, page and line number (line numbers are indicated on the PDF file):

Example 1.

0101.011

= chapter 1, page 1, line 1. The final 1 indicates it is the first unique number and possibly the only one on line 1.

Example 2

1912.192

= chapter 19, page 12, line 19, 2nd unique number on line 19.

I hope you find these instructions helpful and clear. If you have any problems or queries, then please do not hesitate to get in touch. You can contact your production editor or contact Karl Howe, Production Operations Manager (Academic), direct on: khowe@cambridge.org

Thank you.

Karl Howe
Production Operations Manager (Academic)
Cambridge University Press