














Indexes and indexing

‘I propose to bring a bill into Parliament to deprive an author who publishes a book without an index of the privilege of copyright, and, moreover to subject him for his offence to a pecuniary penalty.’

*Lord John Campbell*¹

Schedule of Topics

Introduction	
Fundamental characteristic	
Preparation of the text	
Index Format	
Alternative indexing procedures	
▸ Marking Text	
▸ Concordance files	
▸ Plain Text Files	
Multiple Indexes	
Macro Indexing Aids	
Bundled Macros	
Assigning macros to the keyboard	
Acknowledgements	

The author welcomes comments, suggestions and corrections

1

ex Google search

Probably LORD JOHN CAMPBELL, 1ST BARON CAMPBELL
Scottish statesman, writer and Lord Chancellor of England
(1779 - 1861)

Introduction

In this context the New Shorter Oxford Dictionary defines an index as

‘A list of things in (usu. alphabetical) order; *esp.* a list, usu. at the end of a book, giving the names, topics, etc., mentioned in the book and the places where they occur.’

Merriam-Webster’s defines an index as

‘a list (as of bibliographical information or citations to a body of literature) arranged usu. in alphabetical order of some specified datum (as author, subject, or keyword): as a; a list of items (as topics or names) treated in a printed work that gives for each item the page number where it may be found ...’

In preparing an index the objective is to provide sufficient entries to enable the user to readily identify the subject matter sought.

While WordPerfect only provides two levels of index entry, it is possible to generate multiple indexes to any document. For example, it might be appropriate to have indexes for people, places, and institutions referred in the text. These indexes would be additional to any Table of Contents; Lists of Authorities, Tables, Figures and Diagrams; First Lines (of poems).

Ideally, any document exceeding 4–6 pages should have on the first page a schedule of contents or section headings—this give the reader an overview of the structure or visual rhetoric of the document. Any larger document (exceeding 20–30 pages) should have a Table of Contents and possibly some form of index. All extended document should have an a Table of Contents and an index; they may also have Lists of Authorities, Tables, Figures etc. In all cases page numbers should be included. For some documents index references to numbered paragraphs may be desirable, but are not yet readily achieved with WordPerfect.

If it is intended and expected that the document will be read on screen, hyperlinks should also be considered for all documents. The index generation process [vers. 11–13] make provision for page number hyperlinks to be created automatically for all index entries. Indexing to paragraph numbers must be done manually.

FUNDAMENTAL CHARACTERISTIC

Computer indexing assumes that all entries are CASE-SENSITIVE EXACT MATCHES of the index entry.

Index entries are limited to a maximum of no more than 63 characters per line: each line must end with a HRT (Enter) code.

Where entries with and without an initial capital all have the same connotation, manually combining the entries under the most appropriate form will ensure readers find all forms of the entry word (or phase). For example, if both ‘Book’ and ‘book’ appear in the text and are marked for the index, there will be separate entries for each variation of the word, even when the same connotation is intended but one entry happens to be the first word of

a sentence. Barry MacDonnell's *Unique.wcm*² optionally does this automatically. Nevertheless, some editorial consolidation of the generated index may be necessary to take account of this case-sensitive characteristic [see [example](#)].

PREPARATION OF THE TEXT

Ideally indexing starts with a secure text.

Secure in this context means the text has been fully edited for content and style, and will not be subjected to further additions or amendments. The text of any captions to images, figures, tables and the like have also been established and agreed, and their final location in the text is settled. Similarly, any footnotes or endnotes should be secure.

Tables of Contents or of Authorities, bibliographies and /or bibliographic references, lists of notes and sources should all be excluded from the indexing process. Usually all preliminary pages would also be excluded. Whether or not appendices and any glossary are included will depend on the content, but a judgement must be made.

Caution

*Material to be excluded can be incorporated in comments [Insert>Comment>Create] but **blank lines or pages** must then be inserted in the relevant places to preserve the pagination of the remainder of the text before the index is generated, and these comments restored to text and any spaces then removed after the index is generated.*

Documents assembled from multiple authors present additional problems. While editorial action can achieve textural and stylistic consistency, there may be unwanted font, language, widows and orphans, and other codes in the document that have slipped by unnoticed, and inconsistencies in the placement of formatting codes. Some can be removed or resolved by simple Find & Replace routines such as 'purging all double word spaces from manuscripts'; others can be identified using the macros offered in the [Macro Indexing Aids](#) section.

If a print-ready document to be indexed includes hyphenation and is formatted either as fully justified or as unjustified text, it should be noted that the 'index feature uses the form of hyphenation in the concordance file when it searches the document to be indexed. (Soft hyphens are entered using the hyphen key and hard hyphens are entered using Ctrl+hyphen key. If in doubt, use both forms in duplicate entries in the concordance file.'³ The index generation process will find and record auto-hyphenated words [AutoHyphen EOL] if the concordance file entry is unhyphenated.

² see Macro Indexing Aids below.

³ extracted from the WP Help file

Caution

If documents are to be transferred in PDF format to third parties for printing, particularly to a commercial printer, it should be noted that the bundled WordPerfect PDF generator does not always do a good job of translating some text or image elements such as Advance codes. Alternative PDF generators such as PDF995 or the earlier Acrobat PDF Writer can do a better job. It will pay to test both the pagination accuracy and the quality of the WordPerfect PDF output.

INDEX FORMAT

Always start an index on a new page.

Indexes are commonly formatted as two or three columns, one or two point sizes smaller than the text font. Moderately increased line spacing aids ready recognition of the individual entries. This combination is both economical and effective. Indexes are enhanced by the addition of alpha section headings [[see example](#)].

ALTERNATIVE INDEXING PROCEDURES

For short self-printed documents it is probably simplest to create a list of headings or subjects manually and add page numbers (or paragraph numbers) manually. Alternatively, for electronic documents, hyperlinks from a schedule of topics may suffice.

For longer or extended documents WordPerfect provides three ways of creating indexes:

- marking entries in the text and then using the index generation routine (Tools>Reference>Index);
- using a concordance file, alone or in combination with a marked text; or
- using plain text files created by other means.

Each alternative involves time and thought to achieve an effective result.

A number of macros are available to aid both alternatives.

If the document is a stand alone project of moderate length the 'MARKED' method may be sufficient and appropriate.

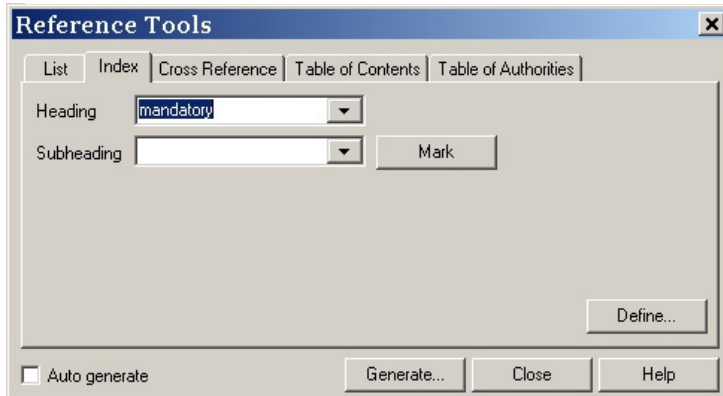
If a simple keyword or phrase only index is appropriate, without secondary subheading entries, a concordance file approach may be preferable.

When a document is part of a series, or one of a number covering inter-related activities or subjects, the concordance file without subheadings has the advantage of being easily edited, expanded, and re-used or saved as a supplementary dictionary. If large or extended documents have discrete sections (or multiple authors), index entries for each completed section can be created and agreed as part of the writing and peer review process, and the section entries aggregated into a final-document concordance file after deleting any duplicated words or phrases.

MARKING TEXT

When you mark an index entry manually, you must designate it as either a heading or a subheading entry. Heading entries are *mandatory*; subheading entries are *optional*.

To mark text, select the word(s), including any relevant font attributes, and then access Tools>Reference>Index (Dialog 1).



Dialog 1: the Index tab selected in this dialog discloses the fields for entering headings and subheadings, and access to the 'Define' options dialog.

If no secondary entries are intended, you can speed up marking Heading entries by selecting the word or phase required and then using the simple macro 'MarkIndexHeadEntry'. This macro finds the highlighted text, places an Index code in the document, and adds the text to the Heading list box. These marked entries will appear in the index after generation.

Double clicking any Index mark in Reveal Codes discloses the selected text that is recorded in the Index code.

For those who prefer keyboarding to mousing about, the macro can be assigned to any convenient keyboard shortcut.

CONCORDANCE FILES

A concordance file is a WordPerfect file which stores all words or phases as alphabetically-sorted CASE-SENSITIVE EXACT MATCHES for the index to your document. Each word or phrase occupies a single line, cannot exceed 63 characters, and ends with HRT. When phrases are included indexes, it may be necessary to provide that the sort process includes the second and subsequent words of any entry. For example, this would ensure that 'animal feed' precedes 'animal husbandry' in the index.


When the index is generated the indexing process sequentially searches the target document for every occurrence of each whole-line entry in the concordance file.

As an example, if the word 'book' appeared in the concordance file, then generating (= compiling) the index finds every occurrence of 'book' in your main document and attaches a page number to those occurrences. Therefore if 'book' appeared on pps. 2, 15, and 16 in your main document then the word 'book' would appear in your index as

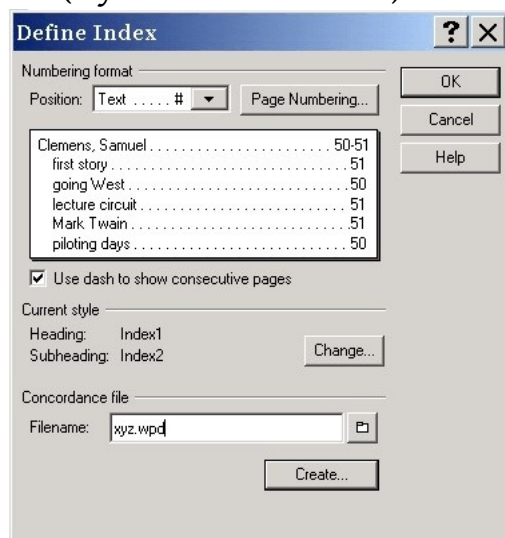
book 2,15,16

or (depending on the page style chosen) as

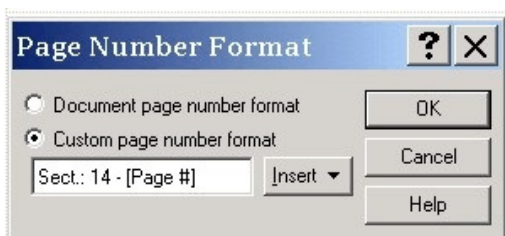
book 2, 15, 16⁴

Similarly, 'Book' may appear only once, say on p. 12. If it has the same connotation as 'book', then the 'Book' entry can be deleted and the page number added to the 'book' entry, e.g. 2, 12, 15, 16 or 2, 12, 15, 16 

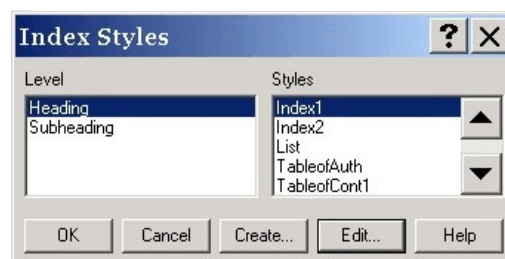
The thing to remember is that the actual index ONLY appears in your main document, usually at the end, but it uses the concordance file to create that index. So, the 'target' file is the main file in which the index is to appear, and the 'source' is either the 'marked' entries in the text and/or the concordance file (if you choose to use one).



Dialog 2: This dialog gives access a variety of numbering formats; to the means of changing the appearance of the headings and subheadings; and to recording the name of the concordance file.



Dialog 2.1: Indicates the way page number formats can be made more informative.



Dialog 2.2: The 'Edit' tab opens a dialog displaying the codes associated with each type of headings: e.g. the relative indentation might be changed; or the font or type attributes changed.

Clearly the various Dialog 2 options give scope for the display of informative and user-friendly indexes.

⁴ The advantage of using the WP Character 4,56 [...] with say 5–8 word spaces between each character over the default continuous line of full points is that the ... characters lead the eye to the number entries but do not 'cut' up the page or column [See the Format>Line>Tab Set for options] see also [Index Example](#).

In suggesting that the font attributes of either main or sub index heading can be changed, bear in mind that font attributed selected with marked text will be included and display in the index. Thus the names of books, boats and aircraft, or binomial names of plants, animals or diseases, traditionally italicised, will be included and displayed in the index.

PLAIN TEXT FILES

Barry MacDonnell has created two macros that produce page number indexes from lists of words or phrases. *Indexlist.wcm* produces the simplest result; *PageLine.wcm* provides page and line numbers details for each entry. Users should read and note the detailed expanded information and particular conditions associated with the use of each macro. They can be found at <http://wptoolbox.com/library/IndexList.html> and <http://wptoolbox.com/library/PageLine.html>.

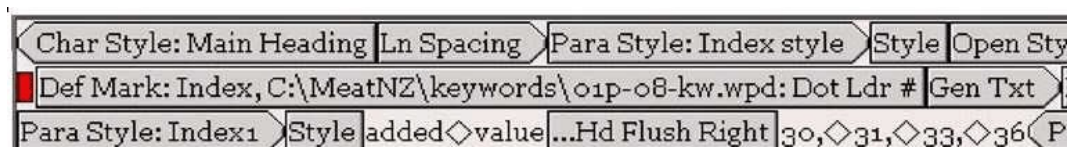
MULTIPLE INDEXES

Multiple indexes⁵ require a little preplanning, and preferably a secure text.

For each index there needs to be a marked text or a concordance file or a combination of both.

If marked text was used for the multiple indexes the marks for previous indexes must be removed and new marks created for the subject of each subsequent index.

However, if multiple indexes are being created, the likelihood is that concordance files will be used, one for each index to be generated.



Dialog 3: The red cursor symbol indicates the start of the index codes. 'Def Mark: Index1' code shows the path and name of the concordance file used, and the dot leader style adopted. 'Gen Txt' code marks the beginning the index text entries. There is a corresponding 'Gen Txt' code immediately after the last index text entry.

To isolate the first index definition and text codes from being processed again, a comment box is created [Insert > Comment > Create] encapsulating the 'Def Mark' and all subsequent codes and text, and must include the end 'Gen Txt' code.

As Dialog 3 shows, generated from a concordance file with only one level of entry, each index entry is a 'mandatory' heading entry treated as a

⁵ I gratefully acknowledge that Sean Mulcahy, a member of the now defunct local WordPerfect User Group, suggested the concept of using comment boxes to isolate individual indexes.

paragraph style which includes the page number style adopted, which in this case is flush right with ... leaders (see [example](#)).

Unless limited for space or print economy issues apply, start each index on a new page.

When the second (and subsequent) indexes are generated, there is no need to compensate for the page(s) occupied by the earlier indexes. The indexes are only going to record the relevant page numbers of the principal text.

When the final index has been generated, return and convert all the previous comment boxes back to text [Insert > Comment > Convert to Text].

If any changes are made to the secure text after indexes have been generated, it will be necessary to delete and regenerate any or all indexes.

Finally, as a quality assurance procedure, random check quite a number of the index entries to ensure that page number do in fact refer to the correct location of the entry. This is particularly important where blank lines or pages have been inserted for any Tables of Contents etc that were isolated from the generating process by being enclosed in a comment box.

Caution

WP X3 may not display the reference tools dialog from the keyboard using Alt + T; C; I when attempting to set up a second index definition. This problem is resolved by editing the Property bar and adding the 'Define Index' and 'Generate' buttons. The various processes then ran smoothly using the buttons.

MACRO INDEXING AIDS

There are a number of macros that significantly reduce the tedium of marking text or creating concordance files. Some are 'zip' bundled with this article; others are identified with their Web address. It may pay to view the macros in 'Edit' mode to see any explanations about the macro features if there is no explanatory text file.

The primary site for these aids in Barry MacDonnell's Toolbox for WordPerfect at <http://wptoolbox.com/library.html> .

His significant personal contribution includes unique.zip — six macros with information and tips to help you index a document. The macro function descriptions below are edited from MacDonnell's supporting document. He has also prepared a paper Creating Indexes.pdf found at <http://wptoolbox.com/library/CreatingIndexes.pdf>

Unique.wcm

This macro creates an alphabetical list of all unique words found in the current document. It creates the list in a separate, 'concordance document,' which is useful in creating an index of the current document. A concordance file created this way should be block-edited to remove unwanted words. In many situations, using the word list created by this macro can be much easier than creating a new word list manually, a word or phrase at a time,

while trying to recall which words should be included.

Caps2End.wcm

This macro locates all words with initial capital letters in a (usually) sorted list and either copies or moves them to the end of the list. It is a companion macro *Unique.wcm*, when *Unique* is used to create a comprehensive list of unique words in a document (for indexing purposes) and the user selects the proper menu choice in *Unique* to create the list while preserving capitalized items. It can, of course, be used standalone on any list of words.

WordFreq.wcm

A macro that produces a word frequency count -- i.e., the number of times each *different* word appears in a document.

RemDupes.wcm

This macro is designed to remove duplicate items from a list of items. It can be used to consolidate several concordance files into one concordance with no duplicate items. It can, of course, be used standalone on any list of words.

IndxHead.wcm

This macro marks the first 64 text characters in all standard and custom paragraph styles (typically, section headings; e.g., Heading 1, Heading 2, MyHeading, etc.) in the document for indexing purposes but check the supporting document for other features.

IndxWord.wcm

Marks the selected word(s) or (if nothing is selected) the word at the cursor location as an index entry (i.e., as an index Heading item).

Another of MacDonnell's indexing aids is his *Alphabetical List Divider.wcm* from the same site, listed in his library under the heading 'ALPHADIV'

Alphabetical List Divider.wcm

This macro divides any alphabetized list such as an index or directory with the appropriate capital letter between each section. ('A,' 'B,' 'C,' etc.) All lines from the cursor position to the end of the document will be processed.

Caution

Some of MacDonnell's macros are quite sophisticated. Potential users should view the bundled supporting document about options and various caveats⁶. Some users may feel more comfortable using traditional WordPerfect procedures or using some of the simpler aids listed below.

MarkIndexHeadEntry.wcm

Used in conjunction with the WordPerfect 9 text selection and the macro assigned to a convenient keyboard shortcut, this is an efficient way of creating single word or phrases entries in a document. For single words, double click the

⁶ Note that some may require prior treatment of the text to achieve expected results, such as having to join words in phrases or proper names with hard spaces to keep them together in the index.

word and invoke the macro. For phases or multiple word entries, double click the first word, hold down the Shift key and select the remaining word(s), and invoke the macro. Reminders about using the macro are contained in an opening prompt.

Caution

If WordPerfect 9 text selection is not used, double clicking a word will include a trailing word space. The macro uses WP 9 text selection and returns the environment to the original state on closure..

MarkEntryConcordance.wcm

This macro captures pre-selected words or phrases and transfers the selected items to a second (pre-opened) document as the basis of a concordance file. The master document and the pre-opened document should be the only documents open.

The caution above applies to this macro.

Hyphen-Dash.wcm

In common with most ASCII-based word processors, hyphens were used to indicate a span of values in dates and other references [1834-42; pp 5-17]. Academic and professional publishing standards would use a short dash rather than a hyphen, so Hyphen_Dash.wcm was created to substitute the WP CharSet value 4,33 [-] for the hyphen.

Another collection of indexing macro aids, Translate.zip, comes with an explanatory document outlining the circumstances and objectives in their creation. Generally these macros are much less sophisticated, but can all be applied to documents after the writing, correcting and peer review functions have been completed. The zip file is found at

<http://wptoolbox.com/otherauthors.html#lindsayrollo>

MisSpelt.wcm

This macro captures as a list all text entries identified by the Spell Checker as 'mis-spelt' or not already added to a custom dictionary. Used unsorted it provides a means of progressively resolving textual queries about spelling, entries having a missing word space; and proper names not recognised by the default and custom dictionaries. As the mis-spelt and miss-typed entries are resolved and deleted, the remaining entries are all potentially candidates for inclusion in an index.

ExtractLC.wcm

Used in conjunction with the lists created by MisSpelt.wcm this macro captures entries commencing with a lower case character and transferred them to a second document; the entries commencing with an upper case character are left in the primary MisSpelt document. The segregation into two lists simplifies editorial scrutiny and assumes that proper names and references to titles of published books, papers and related documents would most likely be correct. Again the remaining entries are all potentially

candidates for inclusion in an index.

CleanUp.wcm

With the best will in the world, users may inadvertently include unwanted characters such as punctuation marks, word spaces, and font attributes and other codes in items marked for an index. If text is being assembled from multiple authors using different word processors, the chances of 'rogue' items increases. Cleanup.wcm removes punctuation and other characters and some codes from lists, sorts the list by the first three words in each line, and deletes case-sensitive duplicate entries. Additional or alternative characters and codes can be easily added.

BUNDLED MACROS

MarkIndexHeadEntry.wcm

MarkEntryConcordance.wcm

Hyphen-Dash.wcm

ASSIGNING MACROS TO THE KEYBOARD

This is the subject of a separate item in PDF format , found at:

http://wptoolbox.com/others/Creating_Custom_Keyboards.PDF

If you are right-handed, chose a shortcut key combination that falls conveniently under the left hand. Using the mouse in the right hand to navigate through the document and to select the entries. Implementing the macro with the left hand speeds up the creation of the index entries.

ACKNOWLEDGEMENTS

I wish to thank Barry MacDonnell for helpful comments and suggestions and for hosting this document on his website.



This document was set 12/16 pt in Gentium Book Basic, an Open Source font available from <http://scripts.sil.org/gentium>

It was chosen for the robust appearance of the serif face and the quality of the PDF files it produces. It provides Basic Latin and Latin-1 Supplement Unicode ranges, plus a selection of the more commonly used extended Latin characters, with miscellaneous diacritical marks, symbols and punctuation. Users should satisfy themselves that it will provide all the characters they need for their particular project.

The current version is licensed under a free/open license and future versions will be similarly licensed.

Lindsay Rollo

August 2008

lrollo@paradise.net.nz

KEYWORD INDEX

A

absorption refrigeration...	...	35
added value...	...	30, 31, 33, 36
aerobic pond...	...	26
Aeromonas...	...	23
ageing...	...	22, 32
air pollution...	...	25
< snip >		
automatic...	24, 33, 34, 36, 53, 60	
automation...	...	24, 33–36
autumn ill-thrift...	...	15, 18

B

bacteria...	16, 20, 23, 24, 26, 30		
barge transport...	54
bating.....	25, 35

bearings...	5, 6
< snip >			
bulletins...	36
bulls...	...	7, 9, 10, 43, 54, 60	
bull-beef...	12
burgers...	31

C

calcium...	6, 31
calf...	7, 10, 57
calves.....	5, 7, 9, 10
calving...	5, 7, 40
Canterbury...	13, 17, 42, 43, 46, 47, 50, 52		
carbon...	23–26, 35
carbon dioxide...	23, 26, 35
carbon dioxide emissions...	26
carcase...	6, 13, 18, 26–28, 30–35, 37, 43,		
carcase processing...	27, 35
carcase splitting...	35

The alpha section headings are styles. These can be edited to provide more or less space above and below the individual characters or change the type face.

The index entries themselves can be given more or less line spacing. Additional line spacing, up to say 1.5 line spaces, can provide greater ease in finding the required entry. However, over 1.5 line spacing the visual rhetoric is rapidly lost. The sample above has 1.3 line spacing.

Where three or more succeeding pages all have a reference to the same word or phrase, WordPerfect inserts a hyphen [-] to show a span of pages. Good publishing practice should use a short dash [–] rather than a hyphen (see examples above). Once the index is generated and before printing, the macro Hyphen_Dash.wcm should be run. This macro substitutes a short dash for the hyphen whenever a numeral is separated from another numeral by a hyphen, such as pps. 23–26, or 15–18 km, or 1835–1847, etc.