

The forindex utilities

Version 0.1

Guido Milanese
guido.milanese@unicatt.it

January 2005

Abstract

Making a good index is a very important part in the process of writing a document, particularly books and manuals. Entering data manually can be a very long process; although a certain amount of data must be entered manually, some tasks can be performed automatically, e.g. an index of geographical names or other trivial tasks. This task can be achieved using the program `doindex`, that prepares a file to be processed by `makeindex`. Another useful feature is to remove all the `\index` entries in a \LaTeX file, obtaining a clean file with no indexing (program `cleanindex`). The programs are written in Snobol4; the only requirement is to install the interpreter. For Windows, a standalone file compiled with Spitbol is also provided, and the programs can be run without the need of an external interpreter.

Contents

1	The programs	2
1.1	The program <code>doindex</code>	2
1.2	The program <code>cleanindex</code>	2
2	Installation	3
2.1	GNU/Linux and other *nix systems	3
2.2	Windows	3
2.3	Windows from source	3
2.4	Cygwin	4
2.5	Macintosh	4
3	Test files	4
4	Bugs and TODO	4
5	Acknowledgements	4
6	Author, copyright, license, disclaimer	5

1 The programs

1.1 The program doindex

The programs `doindex` reads a \LaTeX file, using a list file, and enters index entries in the file according to this list. Previously entered index entries are left unchanged, making it possible to add further indexing to an already indexed file.

The input \LaTeX file may have extension `tex` or `latex`, both uppercase and lowercase (not mixed as in `Tex`).

The list file is meant to contain all the words to be indexed. It must have exactly the extension `wls`. Sub-entries are identified with the separator character `'/'`. See `test.wls` as example:

```
animals
dogs/animals
cats
house/nouns/english/languages
sleeping@sleep
évíta/Italian/foreign words
ça/French/foreign words
drücken/German/foreign words
```

No particular order in this file is required. Some users will prefer alphabetical order, others different orders, so the programs has no requirements concerning order/sort in this file. Entries as `sleeping@sleep` use the standard `makeindex` syntax and are left unchanged. The “logic” of this syntax is opposite to the internal logic of `makeindex`, that is – I think – very clever at the stage of typesetting an index, but not at the stage of designing an index. “A dog is an animal” (`dog/animal` in my syntax) seems to me to be more natural than “Among animals there are dogs” (`animals!dogs` in the `makeindex` syntax). The \LaTeX file produced by `doindex` follows, of course, the `makeindex` conventions.

The original \LaTeX file is left unchanged. A new file will be written, identified by `-ind`. For example, from `file.tex` you will get `file-ind.tex`. Of course, you’ll have to run `makeindex` as usual.

The purpose of the program is similar to what is provided by the program `ixgen` (<http://www.iit.upco.es/~oscar/ixgen/>) written by OSCAR LOPEZ (`oscar@iit.upco.es`), but `forindex` was designed to be a bit more flexible.

1.2 The program cleanindex

The program `cleanindex` removes `\index` sequences from a \LaTeX file. The program can be used e.g. if a user is not happy with the indexing of a file and wants to start it over again.

The input file may have extension `tex` or `latex`, both uppercase and lowercase.

The original file is left unchanged. A new file will be written, identified by `-noind`. For example, from `file.tex` you will get `file-noind.tex`. In this file, lines concerning `makeindex` will be left but commented, in order to avoid an empty

Contents section in the output. You can uncomment the lines as soon as you want to reindex the file again.

2 Installation

2.1 GNU/Linux and other *nix systems

1. Install `snobol4` from <http://www.snobol4.org>. This is Philip Budne's CSNOBOL implementation. You need a `c` compiler to compile the interpreter; it's normally a very quick and easy process.
2. Make sure `snobol4` is in your `PATH` or make a symbolic link.
3. Copy all the files from the source directory in a suitable directory (you do not need the `bat` files, provided for Windows, and can safely remove them).
4. Make executables the scripts (`doindex` and `cleanindex` with no extensions), e.g. `chmod +x doindex`
5. Run the scripts as follows:
 - (a) – to index a text: `./doindex file.tex`
 - (b) If you want to exclude words with accents: `./doindex file.tex --noacc`
 - (c) – to remove `\index` sequences: `./cleanindex file.tex`
 - (d) If the current directory is in your `PATH`, you do not need `./` before the script name.

2.2 Windows

The package offers `exe` files compiled with `Spitbol` (see (<http://www.snobol4.com>)). Make a directory and copy all the file in the `bin/windows` directory. There must be two `*.exe` files and the two `test.*` files.

Run the programs as follows:

– to index a text: `doindex file.tex`

If you want to exclude words with accents: `doindex file.tex -noacc`.

Accents must be encoded using the `latin1` encoding (see the list of `Todo`).

– to remove `\index` sequences: `cleanindex file.tex`

2.3 Windows from source

Basically, follow the same directions given about GNU/Linux, but make sure to use the `bat` files and to install the Windows version of the interpreter. Before using the sources, that are in Unix format, use a script to translate from Unix to Dos-Windows format. If you do not have such a script, open the files with a text editor and save the sources in Windows-Dos format. This can be done reading and saving each file with the DOS `edit` program, with `vim` or any other editor able to deal with different file formats. Do not alter the files if you are

not sure of what you are doing. Please (1) do not use a word processor (as Word or similar) but a simple text editor and (2) make sure to leave the encoding of file acc.inc to ISO-8859-1 or 8859-15, not to plain DOS or Unicode.

2.4 Cygwin

I suggest to follow the same directions given for GNU/Linux, but the EXE files provided for native Windows can be used anyway if preferred.

2.5 Macintosh

Not yet tested (I do not have a Mac right now). It's in the TODO list.

3 Test files

Please test the program on test.tex and test.wls. The produced file will be called test-ind.tex if you use doindex, test-noind.tex if you use cleanindex.

4 Bugs and TODO

The program does not support Unicode files. At this moment, most \LaTeX users are still using latin1, but the situation is rapidly changing.

List of features that I would like to add:

1. Index also included files.
2. Add typographical styles, such as italics for the most important locations of a word.
3. Add support for several indexes (particularly with class memoir)
4. Add an option to generate a rough index for all the words.
5. Add a support to index words listed with regular expressions. E.g. `read*` should index *read*, *reads*, *reading*, *readings*, all under the same heading *read*.
6. Make possible to use another separator for the list file, e.g. a simple blank or other char preferred by user.
7. Test the programs on a Mac.
8. Add Unicode support.

5 Acknowledgements

The program `ixgen` gave me the idea of `forindex`. Many thanks to OSCAR LOPEZ for this very good program.

Some questions sent by CARLO PELLEGRINO (Modena University, Italy) gave me the idea of transforming a very rudimentary script into a general purpose utility. MAURIZIO LORETI (Padua University, Italy) sent me very useful

remarks on the problems of automatical generations of indexes, which I made use of in the introduction to this text.

My warmest thanks to PHIL BUDNE (phil@ultimate.com) for making his excellent CSNOBOL available. Many thanks to the community of Snobol users, particularly to the members of the list snobol4@mercury.dsu.edu, and, among them, to GORDON PETERSON (<http://personal.terabites.com/>), MICHAEL RADOW (mikeradow@yahoo.com), GREGORY L. WHITE (glwhite@netconnect.com.au) and to RAFAL M. SULEJMAN (rafal@engelsinfo.de) whose vim syntax files are a daily blessing.

Thanks to Jim Hefferon [<ftpmaint@alan.smcvt.edu>](mailto:ftpmaint@alan.smcvt.edu) who pointed out that the original name of the package, 4index, was not acceptable due to XML syntax rules.

6 Author, copyright, license, disclaimer

This program is Copyright © 2005
Guido Milanese [<guido.milanese@unicatt.it>](mailto:guido.milanese@unicatt.it)
under the terms of the GNU General Public License.

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

If you do not have a copy of the GNU General Public License write to the Free Software Foundation, Inc., 675 Mass Ave, Cambridge, MA 02139, USA.

If the author of this software was too lazy to include the full GPL text along with the code, you can find it at: <http://www.gnu.org/copyleft/gpl.html>.