

# SCP Help

Alan Reed

November 10, 2017

# Contents

<b>1</b>	<b>Getting Started</b>	<b>1</b>
1.1	What is SCP? . . . . .	1
1.2	Beginner . . . . .	1
<b>2</b>	<b>SCP Projects</b>	<b>2</b>
2.1	An SCP Project . . . . .	2
2.2	Opening an SCP Project . . . . .	2
2.3	Creating a new SCP Project . . . . .	2
2.4	Creating an SCP project from one text file . . . . .	2
2.5	Saving an SCP Project . . . . .	3
<b>3</b>	<b>Text Files</b>	<b>4</b>
3.1	Creating a text file . . . . .	4
3.2	Scanning rules . . . . .	4
3.3	Alphabet . . . . .	4
3.4	References . . . . .	5
3.5	End of Line . . . . .	5
3.6	Join . . . . .	5
3.7	End Hyphen . . . . .	5
<b>4</b>	<b>Sample Projects</b>	<b>6</b>
4.1	Sample Projects . . . . .	6
4.2	Lincoln's Gettysburg address . . . . .	6
4.3	A Tale of Two Cities . . . . .	6
4.4	Sir Gawain and the Green Knight . . . . .	7
<b>5</b>	<b>Key Words</b>	<b>8</b>
5.1	Key Words . . . . .	8
<b>6</b>	<b>Text Processing Commands</b>	<b>9</b>
6.1	Concordance . . . . .	9
6.2	Word List . . . . .	9
6.3	Statistics . . . . .	9

<b>7 Copyright</b>	<b>10</b>
7.1 Text Copyright . . . . .	10
7.2 SCP Copyright and Legals . . . . .	10
<b>8 Futher Reading</b>	<b>11</b>
8.1 Books . . . . .	11
<b>Index</b>	<b>11</b>

# Chapter 1

## Getting Started

### 1.1 What is SCP?

Simple Concordance Program (SCP) is designed help you look through computer readable text files for the occurrence of key words. When each word is found SCP will display the context in which the word is used, and also gives a reference to where that word occurs in the text. SCP will show the vocabulary used by the author and the frequency with which words are used. These can be displayed in alphabetic order, as in a dictionary, or according to their frequency of use. SCP will also produce a word frequency profile and some statistical properties of the texts.

### 1.2 Beginner

As new user of SCP you should first try using an existing project. On the main screen use the File/Open menu item and select *lincoln.scp*. Tap the Open button and you will return to the main screen. On the Concordance Tab, type the word *and* into the Keys box and select the 'word' radio button. Then press the Kwic button. You should see all instances of the use of *and* appear with a line number at the start of each example. Now press the Line button. This shows the lines where the word *and* is used. Replace the word *and* in the Keys box with the letters *th* and select the 'prefix' radio button. Now when you press the Kwic button you get all instances of all words that have *th* as a prefix, so occurrences of *the*, *that* etc appear.

Select the Word List tab and press the 'word list' button. You can now see each word in the vocabulary together with its frequency of occurrence. Un-check the 'Frequencies' check box. Press the 'word list' button again. Now the words are displayed *without* a frequency count. The accompanying drop box allows you to change the layout of the words. Try choosing 'one column' and press the 'word list' button again. Each word of the vocabulary is now displayed on a separate line. You can use the File/Save menu item to write this list to a file, or you can print it using File/Print or use the relevant icon button.

Now select the Statistics tab and press the 'statistics' button. The Word Profile panel shows how words of a given frequency contribute to the whole word count. The Project Statistics shows a few facts about this project, and the Letter Frequencies panel gives counts for each letter used in this project.

Select the menu item View/Project. This shows where the text for the project is located, the alphabet used and other 'scanning rules' that are used to define what SCP recognises as a word. You can define your own projects containing your texts and your scanning rules. Finally, repeat these tasks with the projects *2cities.scp* and *gawain.scp*, they contain extra information called text references which identify the author, page number, and line number.

## Chapter 2

# SCP Projects

### 2.1 An SCP Project

This is a collection of texts, the alphabet used in them, a notation for references, and the vocabulary used by the author. You create an SCP project and save it in a file with an extension of .scp to identify it as an SCP project. You can have many projects which use the same collection of texts but differ only in the scanning rules. The .scp project file is written in the eXtensible Markup Language (XML).

### 2.2 Opening an SCP Project

To open an SCP project click the File Open menu item. All files with the extension .scp will be displayed. Highlight the project you want to open and press the Open button.

### 2.3 Creating a new SCP Project

To create a new SCP project click the File New menu item. You will be presented with a window in which you can select the files to use in your project, the alphabet required, and the type of references used in the text. To choose files for your project highlight one or more files and click 'add files to the project'. A double click of a file adds it to the project. Make all the changes to the alphabet and reference details, or just accept the defaults by pressing the OK button.

SCP defines a word to consist of one or more contiguous letters taken from the alphabet defined here. The order in which words are sorted is based on the order in which letters appear in this alphabet. Normally, capital letters are assumed to be equivalent to small letters, so that a word containing capital letters is counted with its small letter equivalent. However, you could choose to count differently capitalised words separately.

### 2.4 Creating an SCP project from one text file

To create a simple SCP project from a single file of text click the File Open menu item. Now press the 'Files of Type:' drop box and select 'Text Files', these have the extension .txt. Highlight the text file you want to analyse and press the Open button. You will be presented with a window in which you can select the alphabet used and the type of references used in the text. Make all the changes to the alphabet and reference details, or just accept the defaults by pressing the OK button. SCP defines a word to consist of one or more contiguous letters taken from the alphabet defined here. The order in which words are sorted is based on the order in which letters appear in this alphabet. Normally, capital letters are assumed to be equivalent to small letters, so that a word containing capital letters is counted with its small letter equivalent. However, you could choose to count differently capitalised words separately.

## 2.5 Saving an SCP Project

When you create a new SCP project and you click the OK button you will be asked if you wish to save the SCP project. You should normally save your project so that all its characteristics will be available the next time you open it. The file saved has an extension .scp . By default SCP will read through all the text files in a project and retain a copy of each word an author uses together with the frequency of use of each word. This information is saved in the SCP project file and is used in searching for concordances, displaying word lists, and the statistics. Although SCP allows you to avoid saving the vocabulary, you should allow it to do so until you are more familiar with the operation of this program. Failure to save the vocabulary will render some features of SCP unusable.



- 4 *ditto*, add to the current word and terminate the word (French ‘)
- 5 *ditto*, terminate the current word and include in the next word
- 6 treat as a single symbol ‘word’

### 3.4 References

**none** There are no reference symbols in the text, non-blank lines numbers are used.

**type1** References are of the form <X *text*> where < introduces a reference > ends a reference, X is a capital letter A to Z inclusive, (a space **must** follow this letter), and *text* is the text of the reference. NOTE that when <L occurs the *text* **must** consist of numerals, and is assumed to mean a line number. If the symbol following the < is *not* a capital letter all symbols between < and > are ignored. An error report will be generated if your text does not conform to these rules. *A reference must not be split across lines.*

E.G. <A Dickens>  
<P II><L 1>

### 3.5 End of Line

This symbol (normally /) separates lines of text, if it occurs in the middle of a line the line count is increased by one.

### 3.6 Join

This symbol (normally +) causes the program to concatenate the current line with the next one. *The + and all symbols to the right of the + are ignored.* To get the correct line count you should include a / on the next line either immediately or following the first word on the line.

### 3.7 End Hyphen

When this symbol (normally -) occurs *at the end of a line* the program assumes that the current word is continued on the next line.



# Chapter 4

## Sample Projects

### 4.1 Sample Projects

Several sample projects are included in the distribution of SCP. Each comes as one or more text files and as an SCP project file. Open a sample project (using File/Open) and select View/Project to look at the options that were used to create it. Examine the text files of the project and see how the options chosen for, say, the alphabet, would cause words to be selected for the vocabulary. By default a ‘word’ is one or more letters or other symbols taken from an alphabet of your choice, normally the hyphen or apostrophe also treated as if they were letters. The sorting order for words is determined by the order which letters appear in your alphabet and by the codes which are associated with each letter.

### 4.2 Lincoln’s Gettysburg address

This is a sample project which you can use to try out some of the features of SCP. The SCP file *lincoln.scp* contains the project details based on the text file *lincoln.txt* which contains text from Lincoln’s Gettysburg address. The alphabet chosen was the default English:-

```
Capital Letters  ABCDEFGHIJKLMNOPQRSTUVWXYZ-'  
Small Letters   abcdefghijklmnopqrstuvwxyz-'  
Codes          111111111111111111111111111133
```

The codes ensure that - and ’ are ignored when

words are sorted.

The End of Line symbol was the default, namely,

/

No text references have been included.

No Join or End Hyphen symbols have been used.

### 4.3 A Tale of Two Cities

This is a sample project which you can use to try out some of the features of SCP. The SCP file *2cities.scp* contains the project details based on the text file *2cities.txt* which contains text from “A Tale of Two Cities” by Charles Dickens, published by Thomas Nelson and Sons, London, Edinburgh, and New York, 1909.

The alphabet chosen was the default English:-

```
Capital Letters  ABCDEFGHIJKLMNOPQRSTUVWXYZ-'  
Small Letters   abcdefghijklmnopqrstuvwxyz-'  
Codes          111111111111111111111111111133
```

The codes ensure that - and ’ are ignored when words are sorted.

The End of Line symbol was the default, namely,

/

The symbols < and > enclose text references

No Join symbol has been used.

An End Hyphen symbol of - has been used.

## 4.4 Sir Gawain and the Green Knight

This is a sample project which you can use to try out some of the features of SCP. The SCP file *gawain.scp* contains the project details based on the text file *gawain.txt* which contains text from “Sir Gawain and the Green Knight”.

The alphabet chosen was customised as follows:-

```
Capital Letters  ABCDEFGHIJKLMNOPQRSTUVWXYZ-<
Small Letters   abcdefghijklmnopqrstuvwxyz-<
Codes           111111111111111111111111111133
```

The codes ensure that – and < are ignored when words are sorted.

The End of Line symbol was the default, namely,  
/

The symbols [ and ] enclose text references

No Join or End Hyphen symbols have been used.

## Chapter 5

# Key Words

### 5.1 Key Words

The Key Words button allows you to select a portion of the vocabulary used in the texts of the project. You can select by 'point and double-click', by selection, by frequency, by prefix or by suffix, or by the presence of certain letters, or by the length of the words. You can refine your selection in a similar manner. You can use these words when producing a concordance or a word list.

## Chapter 6

# Text Processing Commands

### 6.1 Concordance

An SCP concordance consists of a keyword followed by a series of citations which show how that keyword is used. Alongside the context is a reference to where in the text the word appears. You can choose to have the keyword centred within the context or to display the line in which the word occurs.

You can sort the centralised citations according to the context that appears to the left or right of the keyword.

You can produce a concordance of a chosen word or all words that have a given prefix, suffix, or contain some particular letters. By selecting keywords you can choose a particular subset of the vocabulary used, which can be pre-sorted before a concordance is produced.

The **word profile** shows how the frequency of individual words contributes to the overall frequency of words in the text and can give an indication of an authors style.

**Project statistics** provide a numerical summary of how words are used

**Letter frequencies** shows the frequency of use of the letters and non-letters in the project. You can use the latter to check that your alphabet correctly covers all symbols.

### 6.2 Word List

The wordlist command allows you to display the vocabulary used in the texts of the project. The words can be the whole vocabulary or, by choosing the Key-Words option, you can select only part of it. You can select the order in which the words are sorted and how they are displayed.

### 6.3 Statistics

Here you can see a variety of numerical and tabular views about the project text.

# Chapter 7

## Copyright

### 7.1 Text Copyright

You can only use texts which are out of copyright, or you have permission of the copyright holder or you own the copyright yourself.

### 7.2 SCP Copyright and Legals

Copyright and Licence Information

SCP Simple Concordance Program is owned and licensed exclusively by Alan Reed Copyright (C) 1997-2017 Alan Reed. All rights reserved. Alan Reed grants to you (“Licensee”) a non-exclusive, non-transferable licence to use the SCP binary code version (hereafter, “Binary Software”) without fee. Licensee may distribute the Binary Software to third parties provided that the copyright notice and this statement appear on all copies. Licensee agrees that the copyright notice and this statement will appear on all copies of the software, packaging, and documentation or portions thereof.

ALAN REED MAKES NO REPRESENTATIONS OR WARRANTIES ABOUT THE SUITABILITY OF THE BINARY SOFTWARE, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT. ALAN REED SHALL NOT BE LIABLE FOR ANY DAMAGES SUFFERED BY LICENSEE AS A RESULT OF USING, MODIFYING OR DISTRIBUTING THE BINARY SOFTWARE OR ITS DERIVA-

TIVES.

By downloading, installing, using or copying this Binary Software, Licensee agrees to abide by the intellectual property laws, and all other applicable laws of the U.K., and the terms of this Licence. Ownership of the software shall remain solely in Alan Reed.

Alan Reed shall have the right to terminate this licence immediately by written notice upon Licensee’s breach of, or non-compliance with, any of its terms. Licensee shall be liable for any infringement or damages resulting from Licensee’s failure to abide by the terms of this Licence.

## Chapter 8

# Further Reading

### 8.1 Books

- P. D. Smith, “An Introduction to Text Processing”, The MIT Press, Cambridge, Massachusetts, 1990 ISBN 0-262-19299-3
- A. Reed, “An XML based Alphabet Notation”, Proceedings of the 6th International conference on the Statistical Analysis of Textual Data, St-Malo, March 2002