WordCruncher

Literature Textual Analysis

10/16/2011
UW-Milwaukee
Emily Zegers

OVERVIEW

WordCruncher is a text retrieval and analysis software program that people may download for free onto a personal computer via the program's website on the World Wide Web. According to its' dedicated website, wordcruncher.com, the software program is able to index texts and search them by its vocabulary terms. Additionaly, users may search a variety of languages.

According to Irizarry (1989) the program "consists of two parts, *IndexETC*, which automatically inserts reference code numbers in the text and creates indexed word lists, and *ViewETC,* which manipulates the indexed data."

Unlike a regular print concordance, the software and online platform allows users to not only perform more complex word searches in texts but also add tags, graphics, and hyperlinks to these text documents. WordCruncher also allows users to create analytical reports based on their searches.

EVOLUTION

Concordances take a great deal of time and effort to create. Because of this only the most special of literary texts have concordances compiled for them. This limitation left literary and linguistic scholars with a need for a better way to do research.

When mini-computers arrived on the scene in the 1960's and became more widely used in the 1970's, the compilation of concordances for research in the humanities was one of the earliest applications, according to University of Arizona researcher Mark Olsen (1987).

When the idea for WordCruncher came along in the 1970's the World Wide Web did not exist. Thus WordCruncher has not always been available in the electronic format that it exists in today. Rather it took the form of a CD-ROM.

Researchers at Brigham Young University conceived of the idea for a concordance operating program when the microcomputer became popular for use in business and research institutions (Olsen, 1987). What they created was the Brigham Young University Concordance Program (4.0). Seeing the potential of this program, the Electronic Text Corporation (ETC) further developed and renamed it the name by which it is known today, WordCruncher. In 1987 the CD-ROM sold for $299 (Olsen, 1987), in 1991 it sold for $249 (Madden, 1991), and later in 1997 when upgraded for the Windows platform sold for $249 (Irizarry, 1997).

WordCruncher has evolved since its original CD-ROM format and is freely available from Brigham Young University's WordCruncher website. I found no literature that accounted for the free access to this technology.

FEATURES

Because WordCruncher version 7.1 is freely available, I downloaded it to my personal desktop computer. The download comes with three texts in WordCruncher's "library". A user may download any digital book they wish into their personal library. WordCruncher 7.1 is in beta testing and will eventually come equipped with a large library of texts.

Literature and linguistic researchers use WordCruncher as an analytic tool. The software program provides more than what a concordance can, meaning it does more than provides lists of

words and their location in the literary text. Researchers can construct and search complex word

arguments after which can be analyzed in report format provided by WordCruncher

Searching

As described by the WordCruncher website, there are three types of searches allowed in

WordCruncher: single word, phrases, and wild card searches that include either an asterisk to

truncate words and retrieve various forms of that word and a question mark to retrieve words that

fit the surrounding letters but substitute any letter for the question mark as long as it makes a

word (e.g. re?d would retrieve words such as read, red, reed, reid, and rend.)

When doing a single word or phrase search, a user would double click on any word or phrase in

the text itself or the text search box and the software would retrieve every occurrence of that

word. If the user wants a different word he or she may choose to look at the WordWheel. The

WordWheel offers similar words to the original word searched.

Using the WordWheel is a more intricate process when searching phrases. Since there would be

more than one word, the WordWheel allows the user to add or replace any or all of the words in

the search edit box.

Phrases must have double quotes around them if the user wants to retrieve all occurrences of the

exact phrase. To retrieve a group of words within 7 words of the first word, a user would instead

use single quotes. More complex phrases require custom logic operators between each word.

Wild card searches allow for only single word searching or searching phrases with custom logic.

Book Reports

The Book Report feature is not exactly what a middle school student would think of when assigned to read and report on a book they read. Instead Book Reports provide four specific types of information about the book being searched.

These types of information include general data about the book, such as the number of chapters or number of versus. It provides a list of the typeset found in the book. It provides vocabulary dispersion, meaning, there is a colored vertical graph overview of each word's occurrence patterns within 1000 word blocks. Last, the Book Report feature provides frequency distribution where the frequency of a word is counted within an entire book or selected sections of a book.

Search Reports

The Search Reports provides information on lexicon, word usage, the frequencies of words and phrases, and paired cognates. When searching lexicon with filters, a researcher will get descriptive data on the typeset, word length, and high and low frequency words. Words can be specifically identified by where they first occur or where they occur specifically in a given reference. This would aid a researcher who wants to identify word usage.

The search matrix allows for the identification of frequencies of words and phrases in premade search lists mapped across a set of books.

The WordCruncher website identifies a fourth type of search report, a paired cognate search report, which I am not quite sure of by its description. WordCruncher.com describes this report as useful "when there is an Expansion lexicon for the language. 'If' finds groups of two words that are cognates of the same lexeme and that occur within a given proximity."

I believe that the term cognate refers to two words that have the same historical origin. If so, a linguistic researcher could trace the historical meaning and development of words in texts.

Search Results Reports

Search Results Reports provides three types of reports: Neighborhood Report, Search Vocabulary, and Lowest Level Frequency Distribution.

Neighborhood Report

This report is explained on WordCruncher.com as providing lists of words that collocate with a word or a phrase. Essentially it identifies words that repetitively are found in the same "neighborhood" or in other words are near one another.

Search Vocabulary

The Search Vocabulary Report lists all the words retrieved by a search. It is a more interesting and useful report when applied to wild card searches where word frequencies of complex searches can be mapped into identified sections of text where they were found.

Lowest Level Frequency

I am not sure that I fully understand this feature. The WordCruncher website touts it as being "very useful for finding examples of groups of words." It explains the report as providing the phrase or verse of a text that contains OR'd words.

I think of OR'd words as being strung together by the Boolean operator 'OR'. Thus if I strung together four words with 'OR', this report would identify where in a text most these four words were found together or near one another.

USERS

WordCruncher was created for pedagogical purposes and literary and linguistic research. Before computers and text-analysis software like WordCruncher (there are other programs like it), concordances were the best means by which researches could analyze text. Being that it takes a great deal of time to create a concordance, there were few available. And yet there was a great deal of literature in comparison. Only texts deemed as the most important, such as the Bible, had a concordance composed for it.

With text-analysis software many texts can be analyzed and done so in ways that a concordance couldn't allow. With complex word searches, researchers can develop an understanding of language and texts beyond what they could do with simple concordances.

A pedagogical example is that of Bingham Young professor, Garot Davis. Professor Davis uses WordCruncher in an undergraduate course on Faust. Davis' students are given an assignment to write a paper based upon suggested sophisticated word searches using WordCruncher and the Faust text (Garrett, 1991). He believes that students can derive greater understanding of the literature from the searches enabled by the technology.

CONCLUSION

I believe that literary research has greatly benefited from text-analysis software technology. The type of text analysis that is afforded by the WordCruncher software and other programs like it were not possible thirty years ago. Not only is the research more sophisticated but more literary works can be analyzed.

I provide this endorsement with the disclaimer that it is seemingly important for literary researchers to understand this complex type of program. Their understanding of the software shapes the way they interact with it and generate their data. To crank out data and not necessarily understand how it was derived seems to escape the purpose of research. It is a daunting task in my mind to understand the sophisticated algorithms that generate this data. Furthermore, researchers must evolve with the software technology as it becomes increasingly more sophisticated.

References

Brigham Young University. WordCruncher: electronic text viewer. Retrieved from
    http://www.wordcruncher.com.

Brigham Young University. WordCruncher documentation. Retrieved from
    http://wordcruncher.byu.edu/wordcruncher/documentation.htm

Garrett, N. (1991). Technology in the service of language learning: trends and issues. *The
    Modern Language Journal*, *75*(1), 74-101.

Irizarry, E. (1997). "Interim" review of an "interim" product: "WordCruncher for Windows."
    Computers and the Humanities, *30*(4), 322-324.

Irizarry, E. (1989). Software programs for literary analysis. *Hispania*, *72*(1), 214-216.

Madden, F. (1991) WordCruncher: the riverside Shakespeare. *Computers and the Humanities*,
    25(2/3), 189-192.

Olsen, M. (1987). Textbase for humanities applications: WordCruncher. *Computers and the
    Humanities*, *21*(4), 255-260.

Wikipedia. History of the World Wide Web. (2011, October 14). Retrieved from
    http://en.wikipedia.org/wiki/History_of_the_World_Wide_Web.