
DCG

Document Concordance Generator

Requirements Specification Review

By Chris Blanchard and Eric Brickner

Information Content Representation

Text Document - A file containing the text from which the concordance is generated.

Text(Lines) - An element of the text document. Individual lines are delimited by a carriage return.

Text(Words) - A string of characters within a line. Individual words are delimited by blank spaces.

Status - The status of a word, as determined by the skip-word module.

Coordinates of Appearance - The line number and page number of occurrence of a word within the input text file.

Concordance Structure - The data structure that contains a list of words occurring within the input text file. Associated with each word within the list are that word's coordinates of appearance.

Concordance Document - A file which contains a tabular-format listing of each word within the text document, and its coordinates of appearance.

Functional Partitioning

User Command Parser (UCP) - This module functions to check the specified input and output documents to ascertain whether they are valid, and notifies the user of any errors.

Line/Word Parser Module (LWPM) - This module functions to parse the input document into its constituent words.

Skip-Word Module (SWM) - This module functions to determine if a given word should be omitted from the concordance structure.

Concordance Structure Builder Module (CSBM) - Given a word from the LWPM, this module functions to build the concordance structure.

Output Format Module (OFM) - This module functions to generate the output document, given the concordance structure.

Control Specification

User Command Parser

Control Inputs: User Commands

Control Outputs: UCP Done

Line/Word Parser Module

Control Inputs: UCP Done

Control Outputs: LWPM Done, Check Skip-Word

Skip-Word Module

Control Inputs: Check Skip-Word

Control Outputs: SWM Done

Concordance Structure Builder Module

Control Inputs: LWPM Done

Control Outputs: CSBM Done

Output Format Module

Control Inputs: CSBM Done

Control Outputs: none.

System States

State 1 (User Command Parser) - This is the initial state. If the opening of the input file is a success, then control is handed to State 2 (Line/Word Parser Module). If unsuccessful, then the program terminates.

State 2 (Line/Word Parser Module) - It is necessary to check each word to see if it is to be excluded from the concordance structure. Therefore, as each word is parsed, control is passed over to the Skip-Word Module. Once all the words have been parsed, control is handed over to the Concordance Structure Builder Module.

State 3 (Skip-Word Module) - After each word is checked against the skip-word list, control is handed back to the Line/Word Parser Module.

State 4 (Concordance Structure Builder Module) - Once the concordance

structure has been completely built, control is passed to the Output Format Module.

State 5 (Output Format Module) - When this module has finished generating the Output Document, the DCG terminates.

Events and Actions

Program Initiation

Event: the DCG is invoked with an input and output filenames.

Action: the UCP determines if the files are valid for the DCG.

Parser Go-Ahead

Event: the USP passes the LWPM the input text file stream.

Action: the LWPM begins to parse the input text file.

Check Skip-Word

Action: the LWPM has found a word.

Event: the Skip-Word Module checks if it is a skip-word.

Output Formatting Initiation

Event: the CSBM finishes building the concordance data structure.

Action: the OFM begins to generate the output text file.

Program Termination

Event: the OFM finishes generating the output text file.

Action: the DCG terminates normally.

Performance Bounds

User Command Parser - Determine the status of the text document and the concordance document. Error conditions shall be reported.

Line/Word Parser Module - Parse the entire input text file into individual words. Errors occurring during this process shall be reported.

Skip-Word Module - Identify if a given word is or is not a member of its Skip-Word List.

Concordance Structure Builder Module - Determine the placement of the word within the Concordance Data Structure. Errors in the construction of this data structure shall be reported .

Output Format Module - Format the output text file given a Concordance Data Structure. Errors shall be reported to the user.

Classes of Test / Expected Software Response

User Command Parser

Non-Existent Input File Test - The user will be notified that the specified input file does not exist.

Write-Only Input File Test - The user will be notified that the user's access to the specified input file is write-only.

Already Existing Output File Test - The user will be notified that the specified output file already exists, and the user will be prompted to enter whether to overwrite that file.

Read-Only Output File Test - The user will be notified that the specified output file already exists, and that the user's access to this file is read-only.

Line/Word Parser Module

Line Parse Test - The user will be notified of the number of lines contained in the specified input text file.

Word Parse Test - The user will be notified of number of words contained within each line of the specified input text file.

Skip-Word Module

Skip-Word Test - The user will be notified of the skip-word status of each word within the specified input text file.

Concordance Structure Builder Module

Data Structure Test - The user will be informed of the contents of the entire concordance data structure.

Output Format Module

Format Test - Each word within the input text file (excluding the words on the skip-word list) will appear within the output text file, along with the coordinates of appearance of each word.