"In the scanning process, the program initiates by opening and meticulously examining the file, meticulously documenting the frequency of each word within the encompassing set of documents. This operation mirrors the Map example outlined in our textbook, showcasing its relevance and applicability. Special attention is given to sanitation: all non-alphabetic characters are methodically stripped away, and each word is uniformly converted to lowercase. The outcome is a meticulously curated map, where each entry pairs a word (String) with its respective frequency (Integer) within the scanned documents. This numerical representation vividly illustrates the prevalence of each word in the corpus.

Taking it a step further, a vital component integrated into the process is the stop word filter. Before a word is incorporated into the map, it undergoes scrutiny against a set of predefined stop words. Should a match occur, the word is promptly rejected. This set of stop words is dynamically loaded from a local text file, ensuring adaptability to varying contexts. While readily available online, a tailored stop word file can be effortlessly crafted by iteratively employing a tag cloud generator. These stop words, characterized by their recurrent occurrence, are subsequently appended to the stop words file. The file itself adheres to a standardized format: one word per line, meticulously sorted in alphabetical order, and all in lowercase. Employing a TreeSet or TreeMap facilitates automatic sorting, although manual sorting can be performed through the DOS command line or Shell interface using the provided command."