**Program 595r**

**(Dale-Chall Readability Grade Level)**

**Program Description:** A text file on disk consists of a list of 3,000 familiar words used for the Dale-Chall readability formula. Write a program to read this list of words and calculate the Dale-Chall readability grade level for a given string of text. The program will determine the number of familiar and unfamiliar words, the average sentence length, and then use these metrics to compute the readability score based on the Dale-Chall formula.

Steps to calculate:

1. **Read the Dale-Chall Word List**:
   * Read the words from the data file.
   * Store these words in a data structure that allows for quick lookup (e.g., a Set).
2. **Process the Given Text**:
   * Split the text into sentences to calculate the number of sentences.
     + Use the split function with the regular expression “[.?!]” – i.e., text.split(“[.?!]”);
   * Split the text into words to calculate the number of words.
     + Use the split function with the regular expression “[.?!]” – i.e., text.split(“\\s+”);
     + Clean the words of any non-alphabetic characters and convert to lowercase. This can be done using a function like replaceAll and the regular expression “[^a-zA-Z]” – i.e., clean = word.replaceAll("[^a-zA-Z]", "").toLowerCase();
     + **NOTE:** For a more accurate score, stem words of suffixes like “s”, “es”, “ed”, “ing”, etc.
   * Determine the number of words that are not in the Dale-Chall word list (difficult words).
3. **Calculate Metrics**:
   * Calculate the average sentence length (number of words per sentence).
   * Calculate the percentage of unfamiliar/difficult words.
4. **Compute the Dale-Chall Readability Score**:
   * Use the formula based on the percentage of difficult words (# of difficult words/# of words) and average sentence length (# of words/# of sentences):
   * If the percentage of unfamiliar words is greater than 5%, add 3.6365 to the raw score.
   * Print the calculated readability score and grade.
5. **Calculate the Grade Level using the Raw Score:**
   * **0 – 4.9:** Grades K-4
   * **5.0 - 5.9**: Grades 5-6
   * **6.0 - 6.9**: Grades 7-8
   * **7.0 - 7.9**: Grades 9-10
   * **8.0 - 8.9**: Grades 11-12
   * **9.0 - 9.9**: Grades 13-15 (College)
   * **≥ 10.0**: Grades 16 and above (College Graduate)

**Data Location:** dale-chall-word-list.txt

**Statements Required:** input, output, files, strings manipulation, loop control, decision making, sets

**Sample Output:**

Enter text to calculate the grade level of:

**The quick brown fox jumps over the lazy dog. Many installations require thorough understanding.**

The Dale-Chall readability grade level is: **9.62 (College)**