

Bonus assignment 7: a better word processor

Course ‘Imperative Programming’ (IPC031)

1 Assignment

The naïve version of `read_word` that uses `>>` in the mandatory assignment may consider strange character combinations as a single word. For instance, in the novel “`Alice’s Adventures In Wonderland.txt`”, provided on Brightspace in the mandatory part of the assignment (“`assignment-07-mandatory-files.zip`”) (the red-underlined parts should not belong to a word) the following entries are accepted as words:

- “Wonderland₂.”
- “whatsoever₂.”
- “think₂-.”
- “(for₂.”
- “distance₂-but”
- “Australia₂.”

It is a known hard problem to invent a solution that works for all conceivable texts. For this reason, we restrict the definition of words as follows:

A *word* is a sequence of one or more occurrences of a *letter-character*, followed by zero or more occurrences of a *connector*.

A *connector* is a single hyphen ‘-’ followed by one or more *letter-characters*.

A *letter-character* is:

1. one of the lower-case letters ‘a’ up to and including ‘z’
2. one of the upper-case letters ‘A’ up to and including ‘Z’
3. the apostrophe ‘’

A *garbage-character* is any character that is not a letter-character.

Examples. According to this definition, “waistcoat”, “rabbit-hole”, and “one-two-threes” are individual words, but “distance--but” has to be read as two words. The “`assignment-07-bonus-files.zip`” file on Brightspace contains an additional text file “`hyphen_madness.txt`” to test your function.

Design and implement an improved version of `read_word` that reads words following the above definition of word.

2 Products

As product-to-deliver you only need to upload to Brightspace “`main.cpp`” that you have created with your solution regarding the bonus assignment.

Deadline

Bonus assignment: Monday October 23, 2023, 15:30h