CSE102 – Homework 7
Submission Deadline: December 06, 2016, 11:30pm

In this assignment, like the previous assignment (HW6) you will read two files, tokens.txt and emails.txt (formats and samples of these files are already provided to you in HW6) and build your data structure as tokens with weights (frequencies) and emails with tags (subject and body). After building this data structure you will classify the emails as spam or no-spam as follows:

(1) If a token is found *n* number of times in an email subject and/or body, and *n = weight* of the token, then the email will be classified as spam, otherwise the email will be classified as no-spam.
(2) If any three of the tokens from the complete list of tokens are found in an email subject and/or body then the email will be classified as spam, otherwise the email will be classified as no-spam. Note: In this case, we are only counting if a token is found (once) in the email, and are not comparing it with its weight.

You are required to implement at least (1) to classify the emails and get 100 points. If you also implement (2) then it will give you 20 bonus points that will be added to your grades for this assignment, i.e. a total of 120 points. To get these bonus points you need to implement both (1) and (2). Your program should also check and print errors as described in HW6 requirement 1.

RULES:
1. Obey honor code principles.
2. Read your homework carefully and follow the directives about the I/O format (data file names, file formats, etc.) and submission format strictly. Violating any of these directives will be penalized.
3. Obey coding convention.
4. Your submission should include the following file and NOTHING MORE (no data files, object files, etc):
    HW06_<Firstname>_<Lastname>_<student number>_ classifier.c
    Do NOT compress the files you submit.
5. Do not use non-English characters in any part of your homework (in body, file name, etc.).
6. Deliver the printout of your work until the last submission date.