Course: CS 101

Assignment: Program 5 algorithm

Name: Alexeo Smith

Email: [asd5b@mail.umkc.edu](mailto:asd5b@mail.umkc.edu)

Due : Oct 20th, 2013

Problem:

This program will involve file manipulation. Specifically, we're going to be processing some slightly- messy real-world data and look for patterns.

This data is taken from the Solar Influence Data Center (http://sidc.oma.be/sunspot-data/dailyssn.php), which is maintained by the Royal Observatory of Belgium. You're given a text file containing sunspot observation data going back to 1818. The format is:

Date[YYYYMMDD] Date[YYYY.Fraction] SunspotsObserved

Like many real-world data sets, there are some cases of missing data. Although there is an entry for most days, some days have a number for SunspotsObserved of 999. These cases should be excluded from analysis.

This data is also noisy; that is, there is an underlying pattern (sunspots vary in an 11-year cycle), with a large amount of random 'noise' (day to day variation) on top of it. In order to filter out the noise, we adjust or 'smooth' the data by averaging several values together; if the noise is normally-distributed, this will tend to cancel out the noise. It's not exact—the average error of several points may not be exactly 0—but the average will vary less than individual measurements.

Algorithm:

1. Open the provide data file, parse it and produce a new data file called MONTHLY.TXT that has the monthly average for each month, in the format: Year[YYYY], Month[MM], and Average(floating point value).

Omit any missing data before computing the monthly average

2. From the monthly.txt file, produce smoothed data as follows:

i) For each month, take that month's average, i.e. the six months before it, and the

six months afterward. The smoothed value is (0.5\*first value in the list + 0.5\*last

value in the list + sum of all other values in the list) / 12.

ii) Omit the cases near the very beginning and end of the file that don't have data for

a full 6 months on either side of them

3) Save the data in the same format as for the monthly-data file, in a new file called SMOOTH.TXT. The smoothed data will be in the last column.

4. Produce a graph of the smoothed data via Excel. Based on the graph, estimate by eye when the next peak in sunspot activity will occur and include it in a comment at the beginning of the source file