Design Exercise

UW CSE 190p

Summer 2012

download examples from the calendar

Exercise

- Given a problem description, design a module to solve the problem

1) Specify a set of functions
   - For each function, provide
     - the name of the function
     - a doc string for the function

2) Sketch an implementation of each function
   - In English, describe what the implementation needs to do
   - This should be about 4-5 lines tops per function

A very high-level "pseudocode"

Read the word sentiments from the file
For each word,
    insert the word and its score to a dictionary called scores

Construct the twitter url
Fetch the twitter search results using the url
For each tweet in the response,
    extract the text
    add up the scores for each word in the text
    return the total

Exercise

Problem 1: Text Analysis

Design a module for basic text analysis with the following capabilities:

- Compute the number of words in a file

- Find the 10 most frequent words in a file.

- Find the number of times a given word appears in the file.

Problem 2: Quantitative Analysis

Design a module for basic statistical analysis of files in UWFORMAT with the following capabilities:

- Create an S-T plot: the salinity plotted against the temperature.

- Compute the minimum o2 in a file.

UWFORMAT:
line 0: site, temp, salt, o2
line N: <string>, <float>, <float>, <float>

Changes

Problem 1

- You should ignore *stopwords*. A list of stopwords is provided in a file, one per line.

- I want to see the top *k* words rather than just the top 10.

Problem 2

- UWFORMAT has changed:

  UWFORMAT2:
  line 0: site, date, chl, salt, temp, o2
  line N: <string>, <string>, <float>, <float>, <float>, <float>

- I want to find the average temperature for site "X"