Middle East Technical University | Department of Computer Engineering

CENG 242

Programing Language Concepts

Spring 2011-2012

Homework 2

Due date: 3 April 2012, Tuesday, 23:55

1 Objective

In this homework, you will write a simple text analysis tool using Haskell. Your function analyze will be given a string containing a text and it will return the alphanumeric character frequencies, word frequencies and if there is a parenthesis mismatch in the text.

2 Specifications

Your function analyze will return the following Haskell structure:

```
data StrStat = StrStat { charfreq  :: [(Char,Int)],
                         wordfreq  :: [([Char],Int)],
                         charcount :: Int,
                         wordcount :: Int,
                         parerror  :: Maybe Int }

-- implement the following function
analyze :: [Char] -> StrStat
```

Where fields are explained as:

`charfreq`
: Is a list of character–integer pairs indicating number of occurrences for all alphanumeric characters in the input string. The list should be sorted in increasing order of characters. For example frequencies of `"araba"` is:
`[('a',3),('b',1),('r',1)]`

`wordfreq`
: Is a list of string–integer pairs indicating number of occurrences for all words in the input string. A word is a consecutive sequence of alphanumeric characters. Any punctuation or symbol other than English letters and decimal digits are considered a boundary for a word. A word is always the longest sequence of alphanumeric characters. The list should be sorted in increasing order of strings. For example frequencies of `"ali topu al.al ali al"` is:
`[("al",3),("ali",2),("topu",1)]`

`charcount`
: Is the total number of alphanumeric characters in the input string.

`wordcount`
: Is the total number of words in the input string.

`parerror`
: Is the indicator for if there is a matching error in the input list and place of error. If there is no error the value should be `Nothing` (see definition of `Maybe a` in Haskell Prelude. The valid parentheses are `(,[,{,},],)`. An error occurs if:

  1. A closed parenthesis has no matching opening parenthesis (i.e `[()])`
  2. A closed parenthesis does not match the last opened parenthesis (i.e `[(])`
  3. An opened parenthesis does not have a closed parenthesis at the end of the string.

  In case of one of these errors found the `parerror` is set to `Just n` where `n` is an integer representing the position of the earliest error. After that, parsing of string continues with other fields but result of the parse will contain this value in `parerror` field.

3 Examples

Examples will be updated in this document and/or the newsgroup in the following days.

4 Submission

Submission will be done via COW. You should upload a single Haskell file called "hw2.hs", which includes your source file.

Late submission: At most 3 late days are allowed, but you will lose 10% for each late day. After 3 days, you get 0.

5 Grading

This homework will be graded out of 100.

6 Cheating Policy

People involved in cheating will get 0 from all homeworks and might be reported to the university's disciplinary actions committee.