Skilaverkefni III / Assignment III

StringSet

Þetta verkefni byggir á forritunarverkefni nr. 10 í kafla 11 (bls. 739) í kennslubókinni. Þið eigið að skrifa klasann `StringSet`, hugrænt gagnatag, sem heldur utan um mengi af strengjum (mengi geymir aðeins eitt tilvik af sérhverju staki) og aðalforrit sem prófar klasann. Skilin fyrir klasann `StringSet` eiga að vera í skránni `stringset.h` og útfærslan í skránni `stringset.cpp`. Aðalforritið er í skránni `main.cpp`. Skila þarf einni .zip skrá `stringSet.zip`, sem innihéldur þessar þrjár skrár, í Mooshak. / *This assignment is based on programming project no. 10 in chapter 11 (page 739) in the textbook. You need to write the class* `StringSet`, *an abstract data type, which contains a set of strings (a set only contains one instance of each element) and a main program which tests the class. The interface for the class* `StringSet` *should be in the file* `stringset.h` *and the implementation in the file* `stringset.cpp`. *The main program is in the file* `main.cpp`. *You need to return a single .zip file called stringSet.zip, which contains these three files, in Mooshak.*

Aðalforritið á að lesa tvö skjöl, `doc1.txt` og `doc2.txt`, búa til tvö strengjamengi sem innihalda orðin (strengina) úr þessum skrám og skrifa þau út. Forritið býr síðan til sammengi skjalanna tveggja og skrifar það út. Síðan les forritið inn fyrirspurn úr skránni `query.txt`, býr til samsvarandi mengi og skrifar það út. Því næst skrifar forritið út fjölda orða í fyrirspurninni ásamt hversu mörg þeirra koma fyrir í sammenginu. Að lokum skrifar forritið út upplýsingar um „líkindin" á fyrirspurninni og skjölunum tveimur. Gera má ráð fyrir að allar skrár innihaldi engin greinarmerki. Dæmi um skrár fylgja verkefninu í Myschool. / *The main program reads two documents,* `doc1.txt` *and* `doc2.txt`, *constructs two string sets which contains the words (strings) from these documents, and writes them out. The program then constructs the union of the two documents and writes it out. Thereafter, the program reads in a query from the file* `query.txt` *and writes it out. Then the size of the query is written out along with how many words of the query appear in the union. Finally, the program writes out the similarity between the query and the two documents. You can assume that all the files do not contain any punctuation characters. Example files are included with the project in Myschool.*

Similarity between query Q and document D is computed by:

$$Sim = \frac{|Q \cap D|}{\sqrt{|Q|} * \sqrt{|D|}}$$

Example input/ouput (using the given example files):

```
Doc1: chocolate ice cream and candy bars are my favorite
things to eat fish broccoli not on list
Doc2: I like to eat broccoli and fish along with tofu
brussel sprouts all day long chocolate ice cream are not
good for you
Union: chocolate ice cream and candy bars are my favorite
things to eat fish broccoli not on list I like along with
tofu brussel sprouts all day long good for you
Query: chocolate cream fish good rubbish
Query size: 5
Found in union: 4
Similarity to doc1: 0.325396
Similarity to doc2: 0.373002
```

Nokkur skilyrði varðandi `StringSet` sem þið verðið að uppfylla: / *Several conditions regarding `StringSet` that you need to fulfill:*

1. The interface of the class is in the file `stringset.h`
2. The implementation of the class is in the file `stringset.cpp`
3. The class has a single constructor – a default constructor.
4. The class stores its strings in a `private` vector
5. The class has a `friend` function which overloads the << operator for writing out the contents of a set.
6. The class has a `friend` function which overloads the + operator for constructing the union of two sets.
7. The class has a `friend` function which overloads the * operator for constructing the intersection of two sets.
8. The class has a public function with the signature
   `string at(int i) const;`
   which returns the elment at index i in the vector. This function should be used in all member functions that need to access a given element.
9. The class has at least three other functions!

10. In your implementation, your are NOT allowed to use the Set class (which is part of the Standard Template Library).