CS106 W2020 - Assignment 09

Due: Monday, March 30, 2020, 11:59 PM

Zhoa	Submit

Soundex names:
ZHAO
ZHOU
ZHU
ZEH
ZEE
ZEA
ZAYA
ZIA
ZOU
ZEHE

Soundex Algorithm

The Soundex algorithm allows us to find last names that sound like other last names. For example, Smith sounds like Smythe. This is done by determining a 4 digit soundex code for each name, and if those soundex codes are the same then we say the names sound alike. For example, Harrigan becomes "H625" and Hershenson becomes "H625" so we say Harrigan sounds like Hershenson. Smith and Smythe are both "S530" so we say they sound alike. If you google the internet you will see various algorithms for soundex. In this assignment you will write a function to implement the soundex algorithm and then use that function to find names that sound like other names.

The version of the soundex algorithm we will use is as follows:

1) Retain the first letter. If the name is "Jones" then retain "J". If the name is "Trump" retain the "T".
2) Replace each subsequent letter with a number using the following:
   - a, e, i, o, u, y, h, w → 0
   - b, f, p, v → 1
   - c, g, j, k, q, s, x, z → 2
   - d, t → 3
   - l → 4
   - m, n → 5
   - r → 6
3) At this time we have a long soundex code including zeroes. For example, Harrigan at this point is the String: "H0660205". At this point the length of the original name "Harrigan" and the soundex code "H0660205" are always the same length (length of 8 in this example name). Now,

for any adjacent pairs, you delete the second occurrence. In this example we remove the second "6" and thus "H0660205" becomes "H060205".

4) You then remove the zeroes. "H060205" becomes "H625".
5) If the soundex code has a length less than 4, then you pad it with zeroes on the right. For example, the name "Munn" becomes "M5" after the previous step, and it is then padded with zeroes to the right to have a total length of 4, and thus it becomes "M500".
6) If the soundex code has a length greater than 4, then you truncate it. For example, the name "Delaportilla" becomes "D41634" after step (4) above, then you truncate it to a length of 4, and thus it becomes "M416".
7) The soundex code of length four is returned by the function soundex().

This assignment is broken into two Steps. [14 marks total]

Step 1: [7 MARKS]

You are to write a function called soundex with the following header line:

function soundex(s)

Given a string s of any length, it returns a String of length 4 which is the soundex code.

For example, soundex("Harrigan") returns "H625".

You must follow the steps in the soundex algorithm shown. There are other soundex algorithms available on the internet but you must use the algorithm above.

Step 2: [5 marks]

This is somewhat similar to "ArraySpellChecker" from the demo code. See the screen shot on the previous page.

Do not use a canvas (ie. noCanvas()).

1) Provided to you is a long list of family names in a file "LastNames.txt". Load it in preload() with code like the following:

surNames = loadStrings("/data/LastNames.txt");

2) Allow the user to type in a name. Save this name in a String named something like "lookupName".
3) Determine the soundex code for lookupName (using your function "soundex").
4) Loop over the array of surNames and find any surNames that have the same soundex code as lookupName.
5) Display all the matching surnames.

General Correctness

Assignments that do not run may receive a grade of 0. Even if you don't complete the entire assignment, don't leave it in a broken state. Make sure it runs so we can find ways to give you part marks.

[2 marks] Coding Style and Efficiency

Follow the course coding style for whitespace and comments. Consult the "Code Style Guide" on LEARN. For example:

- Comment your code appropriately. Avoid superfluous comments.
- Correctly and consistently indent your code blocks.
- Use correct inline spacing for variable declaration and assignment.
- Use good line spacing to chunk sections of your code.
- Pay special attention to inline spacing for your conditional statements.
- Use semicolons.
- Use let or const, never var.
- All variables must be declared using let or const. Don't use variables that have not been declared.

One or more marks may be deducted for solutions that have obvious inefficiencies.

- Variables that are declared or assigned, but not used.
- Unnecessarily variables that are duplicates of other variables.
- Unnecessarily repeating the same code in multiple places.
- Too many "magic numbers": the same number appears in many places indicating a variable should have been used instead.

Restrictions

In general, you may not use any functions, libraries, or statements not covered in lecture or labs unless not specifically exempted in a post by a TA or instructor on this assignment discussion board.

Submitting

Zip your sketch folder (CS106_A09), and submit it the correct assignment dropbox. Consult "How to Submit" on Learn for more information on how to create a ZIP.

It is your responsibility to submit to the correct dropbox with the correct files before the deadline. Otherwise you will receive a mark of 0.