

CP #8: Soundex Indexing

The Soundex Index System was developed so that similar sounding names, or names with similar spelling could be encoded for easy retrieval. It has been used by the U.S. Bureau of the Census, and some States use it to help encode your driver's license number. Your task is to read a sequence of names, one at a time and one per line, compute the corresponding soundex code, and write to the output file the name and its soundex code (one line of output per name).

Names will contain from 1 to 20 upper case, alphabetic characters (ASCII values 65 thru 90, inclusive). Names shorter than 20 characters will NOT be padded with blanks. Thus a name will consist of upper case letters only.

How to generate the Soundex Code:

A Soundex Code always consists of a letter followed by three digits. Here are the rules for soundex encoding:

1. The first letter of a name appears as the first and only letter in the soundex code.
2. The letters A, E, I, O, U, Y, W and H are never encoded, but do break successive code sequences (see next rule).
3. All other letters are encoded EXCEPT when they immediately follow a letter (including the first letter) that would be encoded with the same code digit.
4. The soundex code guide is:

Code	Key Letters and Equivalents
1	B, P, F, V
2	C, S, K, G, J, Q, X, Z
3	D, T
4	L
5	M, N
6	R

5. Trailing zeros are appended to short codes so all names are encoded with a letter followed by three digits.
6. Longer codes are truncated after the third digit.

Input Format

The input file contains a list of names, one per line. Each name will not exceed 20 characters, and you may assume that only upper case letters will be used. Your program should continue to read names until the end of the file is detected.

Output Format

The output written to the file should consist of a column of names and a column of their corresponding soundex codes. Write the headings 'NAME' and 'SOUNDEX CODE' in the first line of the output file in columns 10 and 35, respectively. After the heading line, the names and soundex codes should be written (one pair per line) with the name starting in column 10 and the soundex code beginning in column 35. The comment 'END OF OUTPUT' should appear at the end of the output file on the line immediately after the last name. This comment should be written starting in column 20.

Sample Input 0

LEE
KUHNE
EBELL
EBELSON
SCHAEFER
SCHAAK

Sample Output 0

LEE;L000
KUHNE;K500
EBELL;E140
EBELSON;E142
SCHAEFER;S160
SCHAAK;S200