

Senior Division**The Soundex Indexing System**

(rules adapted from the National Archives and Records Administration)

PROBLEM: Interested in your genealogy? Trying to find a long lost friend? To search large databases such as the US Census records, you need to use the Soundex Indexing System. The soundex is a coded surname (last name) index based on the way a surname sounds rather than the way it is spelled. Surnames that sound the same, but are spelled differently, like SMITH and SMYTH, have the same code and are filed together. The soundex coding system was developed so that you can find a surname even though it may have been recorded under various spellings.

To search for a particular surname, you must first work out its code. Every soundex code consists of a letter and three numbers, such as W252. The letter is always the first letter of the surname. Now, disregard the letters A, E, I, O, U, H, W, and Y from the remaining letters of the name. The numbers are assigned to the remaining letters of the surname according to the soundex guide shown below. Zeroes are added at the end if necessary to produce a four-character code. Additional letters are disregarded. Examples:

WASHINGTON is coded W252 (W, 2 for the S, 5 for the N, 2 for the G, remaining letters disregarded).

LEE is coded L000 (L, 000 added).

Number	Represents the Letters
1	B, F, P, V
2	C, G, J, K, Q, S, X, Z
3	D, T
4	L
5	M, N
6	R

Other soundex rules are as follows:

1. Names With Double Letters

If the surname has any double letters, they should be treated as one letter. For example:

GUTIERREZ is coded G362 (G, 3 for the T, 6 for the first R, second R ignored, 2 for the Z).

2. Names with Letters Side-by-Side that have the Same Soundex Code Number

If the surname has different letters side-by-side that have the same number in the soundex coding guide, they should be treated as one letter. Examples:

PFISTER is coded as P236 (P, F ignored, 2 for the S, 3 for the T, 6 for the R).

JACKSON is coded as J250 (J, 2 for the C, K ignored, S ignored, 5 for the N, 0 added).

TYMCZAK is coded as T522 (T, 5 for the M, 2 for the C, Z ignored, 2 for the K). Since the vowel "A" separates the Z and K, the K is coded.

3. Names with Prefixes

If a surname has a prefix, such as VAN, CON, DE, DI, LA, or LE separated from the surname by a space, code the name with and without the prefix because the surname might be listed under either code. Note, however, that Mc and Mac are **not** considered prefixes. For example:

VAN DEUSEN is coded two ways:

V532 (V, 5 for N, 3 for D, 2 for S) and D250 (D, 2 for the S, 5 for the N, 0 added).

4. Consonant Separators

If a vowel (A, E, I, O, U) separates two consonants that have the same soundex code, the consonant to the right of the vowel is coded. Example:

TYMCZAK is coded as T522 (T, 5 for the M, 2 for the C, Z ignored (see "Side-by-Side" rule above), 2 for the K). Since the vowel "A" separates the Z and K, the K is coded.

If "H" or "W" separate two consonants that have the same soundex code, the consonant to the right of the vowel is **not** coded. Example:

ASHCRAFT is coded A261 (A, 2 for the S, C ignored, 6 for the R, 1 for the F). It is not coded A226.

INPUT: There will be 5 lines of input. Each line will contain one string with all upper case letters.

OUTPUT: For each input print the 4- character soundex code for that string.

SAMPLE INPUT

1. WASHINGTON
2. LEE
3. VAN DEUSEN

SAMPLE OUTPUT

1. W252
2. L000
3. V532, D250