

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.

Basic Soundex Coding Rule

Every soundex code consists of a letter and three numbers, such as W-252. The letter is always the first letter of the surname. 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 W-252 (W, 2 for the S, 5 for the N, 2 for the G, remaining letters disregarded).

W	A	S	H	I	N	G	T	O	N
W		2			5				2

Lee is coded L-000 (L, 000 added).

L	E	E	
L	0	0	0

Soundex Coding Guide

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

Disregard the letters A, E, I, O, U, H, W, and Y.

Additional Soundex Coding Rules

Names With Double Letters

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

Gutierrez is coded G-362 (G, 3 for the T, 6 for the first R, second R ignored, 2 for the Z).

G	U	T	I	E	R	R	E	Z
G		3			6			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 P-236 (P, F ignored, 2 for the S, 3 for the T, 6 for the R).

P	F	I	S	T	E	R
P	2	3	6			

Jackson is coded as J-250 (J, 2 for the C, K ignored, S ignored, 5 for the N, 0 added).

J	A	C	K	S	O	N
J	2	5	0			

Tymczak is coded as T-522 (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.

T	Y	M	C	Z	A	K
T	5	2	2			

Names with Prefixes

If a surname has a prefix, such as Van, Con, De, Di, La, or Le, code both with and without the prefix because the surname might be listed under either code. Note, however, that Mc and Mac are **not** considered prefixes. Example:

VanDeusen might be coded two ways:

V-532 (V, 5 for N, 3 for D, 2 for S)

V	A	N	D	E	U	S	E	N
V	5	3	2					

or D-250 (D, 2 for the S, 5 for the N, 0 added).

D	E	U	S	E	N
D	2	5	0		

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 T-522 (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.

T	Y	M	C	Z	A	K
T	5	2	2			

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 A-261 (A, 2 for the S, C ignored, 6 for the R, 1 for the F). It is not coded A-226.

A	S	H	C	R	A	F	T
A	2	6	1				