



Genealogical Computing Group

Special Interest Group of the
New Zealand Society of Genealogists
Southland Branch



Ever wondered why, after turning on Soundex on a web search, you get unexpected results and assumed it is because American cannot pronounce the name you are searching for correctly. Well after researching Soundex, so that I could implement it in the Societies new local databases, I can no longer blame American pronunciation. The system only looks at the first letter and the next three unique consonants, which are grouped into six groups. Its wide use is because of its simplicity, and it being built into most computer languages supported by Web Servers. Other more complex systems have been developed, but Soundex is the most popular.

Soundex is a method for indexing names by sound, as pronounced in English, developed by Robert Russell and Margaret Odell in 1918. The goal is for names with the same pronunciation to be encoded to the same way so that they can be matched despite minor differences in spelling.

Rules:- The Soundex code for a name consists of a letter followed by three digits: the letter is the first letter of the name, and the digits encode the remaining consonants. Similar sounding consonants share the same digit. Vowels and the letters H, W and Y are not coded themselves, except as the first letter.

Soundex Coding :-

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

- Collapse *adjacent* identical digits into a single digit of that value.
- Return the starting letter and the first three remaining digits. If needed, append zeroes to make it a letter and three digits.

Using this algorithm, both "Robert" and "Rupert" are coded as "R163" while "Rubin" is "R150".

Note that surname prefixes such as van, Von, Di, de, le, D', dela, or du are sometimes disregarded in alphabetizing and in coding.