Phonetic algorithm

From Wikipedia, the free encyclopedia

> This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(August 2009)*

A **phonetic algorithm** is an algorithm for indexing of words by their pronunciation. Most phonetic algorithms were developed for use with the English language; consequently, applying the rules to words in other languages might not give a meaningful result.

They are necessarily complex algorithms with many rules and exceptions, because English spelling and pronunciation is complicated by historical changes in pronunciation and words borrowed from many languages.

Among the best-known phonetic algorithms are:

- Soundex, which was developed to encode surnames for use in censuses. Soundex codes are four-character strings composed of a single letter followed by three numbers.
- Daitch–Mokotoff Soundex, which is a refinement of Soundex designed to better match surnames of Slavic and Germanic origin. Daitch–Mokotoff Soundex codes are strings composed of six numeric digits.
- Kölner Phonetik: This is similar to Soundex, but more suitable for German words.
- Metaphone, Double Metaphone, and Metaphone 3 ⧉ which are suitable for use with most English words, not just names. Metaphone algorithms are the basis for many popular spell checkers.
- New York State Identification and Intelligence System (NYSIIS), which maps similar phonemes to the same letter. The result is a string that can be pronounced by the reader without decoding.
- Match Rating Approach developed by Western Airlines in 1977 - this algorithm has an encoding and range comparison technique.
- Caverphone, created to assist in data matching between late 19th century and early 20th century electoral rolls, optimized for accents present in parts of New Zealand.

Common uses [edit]

- Spell checkers can often contain phonetic algorithms. The Metaphone algorithm, for example, can take an incorrectly spelled word and create a code. The code is then looked up in directory for words with the same or similar Metaphone. Words that have the same or similar Metaphone become possible alternative spellings.
- Search functionality will often use phonetic algorithms to find results that don't match exactly the term(s) used in the search. Searching for names can be difficult as there are often multiple alternative spellings for names. An example is the name Claire. It has two alternatives, Clare/Clair, which are both pronounced the same. Searching for one spelling wouldn't show results for the two others. Using Soundex all three variations produce the same Soundex code, C460. By searching names based on the Soundex code all three variations will be returned.

See also [edit]

- Approximate string matching
- Hamming distance
- Levenshtein distance
- Damerau–Levenshtein distance

References [edit]

- Black, Paul E. "phonetic coding" ⧉. *Dictionary of Algorithms and Data Structures*. NIST.

External links   [edit]

- Algorithm for converting words to phonemes ⧉ and back.
- StringMetric project ⧉ a Scala library of phonetic algorithms.
- clj-fuzzy project ⧉ a Clojure library of phonetic algorithms.
- SoundexBR ⧉ library of phonetic algorithm implemented in R.

This algorithms or data structures-related article is a stub. You can help Wikipedia by expanding it.

['stʌb] *This phonetics article is a stub. You can help Wikipedia by expanding it.*