The phonics package provides implementations of several phonetic algorithms. Phonetic algorithms are used to encode a string based on how it is pronounced (Zobel and Dart 1996). The resultant code should provide functional matching between similarly pronounced names. For instance, “Robert” and “Rupert” both have the Soundex value of “R163,” suggesting they are pronounced almost identically. Because of pronunciation differences around the world, even across English, many different algorithms exist and serve different needs and populations.

The algorithms are typically used for name encoding and indexing, record linking between unrelated databases, and spellchecking. In addition, they can be used as a proxy for string distance measurements. Included in this package are Soundex, Metaphone, and many others developed over the years, including published variants.

Acknowledgements

The author thanks Oliver Keyes for his contributions and improvements to the C++ implementations within this package.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 (Towns et al. 2014). In particular, it used the Comet system at the San Diego Supercomputing Center (SDSC) (Moore et al. 2014; S. M. Strande et al. 2017) through allocations TG-DBS170012 and TG-ASC150024.

References

