

# Phonetic Algorithms in R

James P. Howard, II<sup>1</sup>

DOI: [10.21105/joss.00480](https://doi.org/10.21105/joss.00480)

<sup>1</sup> The Johns Hopkins University Applied Physics Laboratory

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

**Submitted:** 15 November 2017

**Published:** 14 February 2018

## Licence

Authors of JOSS papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

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 (Townsend 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

Moore, Richard L., Chaitan Baru, Diane Baxter, Geoffrey C. Fox, Amit Majumdar, Phillip Papadopoulos, Wayne Pfeiffer, et al. 2014. “Gateways to Discovery: Cyberinfrastructure for the Long Tail of Science.” In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, 39:1–39:8. XSEDE ’14. New York, NY, USA: ACM. <https://doi.org/10.1145/2616498.2616540>.

Strande, Shawn M., Haisong Cai, Trevor Cooper, Karen Flammer, Christopher Irving, Gregor von Laszewski, Amit Majumdar, et al. 2017. “Comet: Tales from the Long Tail: Two Years in and 10,000 Users Later.” In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*, 38:1–38:7. PEARC17. New York, NY, USA: ACM. <https://doi.org/10.1145/3093338.3093383>.

Townsend, John, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, et al. 2014. “XSEDE: Accelerating Scientific Discovery.” *Computing in Science & Engineering* 16 (5). IEEE:62–74.

Zobel, Justin, and Philip Dart. 1996. "Phonetic String Matching: Lessons from Information Retrieval." In *Proceedings of the 19th Annual International Acm Sigir Conference on Research and Development in Information Retrieval*, 166–72. ACM.