NAME

Text::Soundex - Implementation of the Soundex Algorithm as Described by Knuth

SYNOPSIS

```
  use Text::Soundex;

  $code = soundex $string;            # get soundex code for a string
  @codes = soundex @list;             # get list of codes for list of
strings

  # set value to be returned for strings without soundex code

  $soundex_nocode = 'Z000';
```

DESCRIPTION

This module implements the soundex algorithm as described by Donald Knuth in Volume 3 of **The Art of Computer Programming**. The algorithm is intended to hash words (in particular surnames) into a small space using a simple model which approximates the sound of the word when spoken by an English speaker. Each word is reduced to a four character string, the first character being an upper case letter and the remaining three being digits.

If there is no soundex code representation for a string then the value of `$soundex_nocode` is returned. This is initially set to `undef`, but many people seem to prefer an *unlikely* value like `Z000` (how unlikely this is depends on the data set being dealt with.) Any value can be assigned to `$soundex_nocode`.

In scalar context `soundex` returns the soundex code of its first argument, and in list context a list is returned in which each element is the soundex code for the corresponding argument passed to `soundex` e.g.

```
  @codes = soundex qw(Mike Stok);
```

leaves `@codes` containing `('M200', 'S320')`.

EXAMPLES

Knuth's examples of various names and the soundex codes they map to are listed below:

```
  Euler, Ellery -> E460
  Gauss, Ghosh -> G200
  Hilbert, Heilbronn -> H416
  Knuth, Kant -> K530
  Lloyd, Ladd -> L300
  Lukasiewicz, Lissajous -> L222
```

so:

```
  $code = soundex 'Knuth';              # $code contains 'K530'
  @list = soundex qw(Lloyd Gauss); # @list contains 'L300', 'G200'
```

LIMITATIONS

As the soundex algorithm was originally used a **long** time ago in the US it considers only the English alphabet and pronunciation.

As it is mapping a large space (arbitrary length strings) onto a small space (single letter plus 3 digits) no inference can be made about the similarity of two strings which end up with the same soundex code. For example, both `Hilbert` and `Heilbronn` end up with a soundex code of `H416`.

AUTHOR

This code was implemented by Mike Stok (`stok@cybercom.net`) from the description given by Knuth. Ian Phillipps (`ian@pipex.net`) and Rich Pinder (`rpinder@hsc.usc.edu`) supplied ideas and spotted mistakes.