https://miro.com/app/board/o9J_lO9EAtw=/

Dictionaries + Tolerant Retrieval (Chapter 3)

Exercise 3/1

a) Find **two different words** of the **same** soundex code.

b)Find **two phonetically similar words** of **different** soundex codes.

cymbal != symbol

Algorithm 1 (Soundex Code)
Transformation of a string to a 4-character soundex code

1. *Keep the first character*
2. *Rewrite* $\{A, E, I, O, U, H, W, Y\}$ *to* 0
3. *Rewrite characters*
   (a) $\{B, F, P, V\}$ *to* 1
   (b) $\{C, G, J, K, Q, S, X, Z\}$ *to* 2
   (c) $\{D, T\}$ *to* 3
   (d) $\{L\}$ *to* 4
   (e) $\{M, N\}$ *to* 5
   (f) $\{R\}$ *to* 6
4. *Remove duplicities*
5. *Remove zeros*
6. *Change to length 4 (truncate or add trailing zeros)*

SWORD

S0063 --> S063 --> S63 --> S630

short

FAX = F*CK F200

Exercise 3/2

Write elements in a dictionary of the permuterm index generated by the term:

mama $

mama$, ama$m, mama, amam, $mama

Algorithm 2 (Querying in Permuterm Index)
For query q, find keys according to the following scheme:

- *for* $q = X$, *find keys in the form* $X\$$
- *for* $q = X*$, *find keys in the form* $\$X*$
- *for* $q = *X$, *find keys in the form* $X\$*$ *X$ -->X$*
- *for* $q = *X*$, *find keys in the form* $X*$
- *for* $q = X*Y$, *find keys in the form* $Y\$X*$

https://nlp.stanford.edu/IR-book/html/htmledition/permuterm-indexes-1.html

economic*
search*

economics, economical,..

Exercise 3/3

Which keys are usable for finding the term

s*ng

in a permuterm wildcard index?

s*ng$ --> ng$s*

searching

sing sang sung

song

spring

string