

Tries with frequencies WYC-2017

Presentation topics

Trie definition

Trie with frequencies definition

Contingency tables and mosaic plots

Sequence mining

Applications to text analysis

ENR with N-gram models

Grammar creation

Word frequencies for POS

Frequent phrases

Phrase completion

Convert time series to sequences of symbols

Then mine and predict with tries with frequencies

Image recognition with tries

Does not work that well

But illustrates a principle

Imagine we know the paths of handwriting...

Route prediction and finding

And subsequent combination with, say, Decision Trees

See my Spatial Method Objects talk at WTC-2012

Real life example with patient throughput

Comparison with RNN

MSE discussions

Infixes in a dictionary

Most common prefixes in a given collection of texts

Trie merging is handy here

Things to mention or emphasize

The first version of the package

Implementations

WL

R

Java

The Java interface package

MosaicPlot is made with tries with frequencies

Naive Bayesian Classifiers can be made with tries with frequencies