

## Preprocessing

- String rawSlice
- String[] strList
- String[] varList

- 
- + preprocess():String
  - + processNumber(String):String
  - + processStr(String):String
  - + processVar(String):String
  - + processType(String):String

## WordTokenDict

- List idx2word
- Map word2idx
- String unkToken

- 
- + addWord(String):String
  - + wtoi(String):int
  - + itow(int):String
  - + save(String):void
  - + load(String):WordTokenDict

## Tokenizer

- TokenDict tokenDict
- int freqGt

- 
- + buildDict(TextDataset):WordTokenDict
  - + encode(String):int[]
  - + decode(int[]):String
  - + collateFn(List):(Tensor,Tensor,Tensor)

