Mallot

Instances- 4 fields

Data: primary object (input)

Name: for the user

Target/Label: what type of output

Source: often empty; what the instance originally look like

Data

* String-> creates token sequence (wrapper for a string) -> feature sequence -> feature vector maps int to double

File Iterator

* All concatenated into one file string

CsvIterator

* Can use to parse tab delimited files (regex)

Instance Pipelines

* Transforms the data

Overall

* Use pipeList to have an ordered sequence of pipes

Classifying

* Can train classifier objects with interface Optimizer

HMM and CRF