

hw1 report

Han Zhang (hanz)

1. machine learning technique: HMM, graphical model
2. NLP technique: NER, chunking,
3. no marked up text
4. gene tag file
5. no rule sets used
6. gene tag is the biological data we interact with
7. Lingpipe is the relevant resources we used for development
8. CAS got text line-by-line from input file and do NER and then get chunks for Gene name. Finally we written out gene by gene to output file.