

Bi-mothly Report

FreeNeb

Ziwei Bai

summary

- Crawl data
 - Crawl html data
 - Parse html data
 - buy data
- Support zy & my
- CDSSM train data processing

summary

- Crawl data
 - Crawl html data
 - Parse html data
 - buy data
- Support zy & my
- CDSSM train data processing

Crawl data

- News
 - Financial 18G
 - Non-financial 79G
- QA
 - Financial 16G
 - Non-financial 3.9G
- BBS
 - Financial 2G
- Other
 - Non-financial 18G

summary

- Crawl data
 - Crawl html data
 - Parse html data
 - Buy data
- Support zy & my
- CDSSM train data processing

Support zy & my

- Thesaurus
 - Jieba
 - Sogou
 - Stock thesaurus
- Decode
 - Sogou Thesaurus
- Label
- Phonetic

summary

- Crawl data
 - Crawl html data
 - Parse html data
 - Buy data
- Support zy & my
- CDSSM train data processing

CDSSM train data processing

- Word2vec
- Cut words
- Split the traindata