没有引进词典的

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Jieba | THULAC | LTP | NLPIR |
| 时间/s | 13.454736065864562 | 18.879502367973327 | 58.370395636558534 | 14.536421036720276 |
| 正确率 | 0.8126232741617357 | 0.665680473372781 | 0.748422712933754 | 0.5867850098619329 |
| 召回率 | 0.6399747793190416 | 0.5252206809583859 | 0.5983606557377049 | 0.4665825977301387 |
| F值 | 0.7112824106517169 | 0.5837421163279608 | 0.6650315346881569 | 0.5192711983181499 |

引进词典的

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Jieba | THULAC | LTP | NLPIR |
| 时间/s | 13.325211834907531 | 19.681711983680724 | 57.69990231990814 | 14.342047524452209 |
| 正确率 | 0.878698224852071 | 0.7534516765285996 | 0.7421135646687698 | 0.796844181459566 |
| 召回率 | 0.692938209331652 | 0.6008827238335435 | 0.5933165195460278 | 0.621689785624212 |
| F值 | 0.770147161878066 | 0.6678346180798879 | 0.6594253679046953 | 0.690960056061668 |

高频药材的频次分析

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 人工分词 | | jieba引进词表的分词 | | thulac引进词表的分词 | | pynlpir引进词表提取的分词 | | plt引进入词表的分词 | |
| 枸杞 | 0.0812 | 枸杞 | 0.0828 | 枸杞 | 0.0757 | 枸杞 | 0.0835 | 枸杞 | 0.0804 |
| 山药 | 0.0749 | 山药 | 0.0749 | 山药 | 0.0686 | 山药 | 0.0749 | 山药 | 0.0733 |
| 山楂 | 0.0583 | 山楂 | 0.0607 | 山楂 | 0.0575 | 山楂 | 0.0607 | 山楂 | 0.0591 |
| 人参 | 0.0567 | 人参 | 0.0583 | 人参 | 0.0504 | 人参 | 0.0583 | 人参 | 0.0583 |
| 茯苓 | 0.0354 | 茯苓 | 0.0362 | 菊花 | 0.0339 | 菊花 | 0.0370 | 菊花 | 0.0362 |
| 菊花 | 0.0354 | 菊花 | 0.0362 | 生姜 | 0.0331 | 茯苓 | 0.0362 | 葛根 | 0.0347 |
| 葛根 | 0.0347 | 葛根 | 0.0354 | 大枣 | 0.0323 | 生姜 | 0.0339 | 生姜 | 0.0331 |
| 大枣 | 0.0331 | 大枣 | 0.0347 | 葛根 | 0.0323 | 葛根 | 0.0339 | 大枣 | 0.0331 |
| 甘草 | 0.0315 | 甘草 | 0.0339 | 甘草 | 0.0299 | 甘草 | 0.0331 | 甘草 | 0.0299 |
| 生姜 | 0.0315 | 生姜 | 0.0331 | 茯苓 | 0.0291 | 大枣 | 0.0283 | 茯苓 | 0.0299 |