# 多关键词查询

实现两种方式：依存句法分析和根据业务逻辑自定义判断和解析逻辑。

## 依存句法分析

### 判断及解析逻辑

依存句法分析不能处理省略关键词的情况。

根据依存句法分析，判断词与词之间的依赖关系，据此解析查询语句。

尝试使用了三种依存句法分析工具，分别为deepnlp、huannlp和哈工大语言云(LTP)。三种工具分别使用不同的方法实现依存句法分析，其中deepnlp利用深度学习的方法，Huannlp利用crf方法，LTP将依存分析问题归结为在一个有向图中寻找最大生成树（Maximum Spanning Tree）的问题。在依存句法分析模块中，LTP分别实现了

* 一阶解码(1o)
* 二阶利用子孙信息解码(2o-sib)
* 二阶利用子孙和父子信息(2o-carreras)

### 实验结果

选取20条样例，分别用三种句法分析工具进行解析。样例如下。

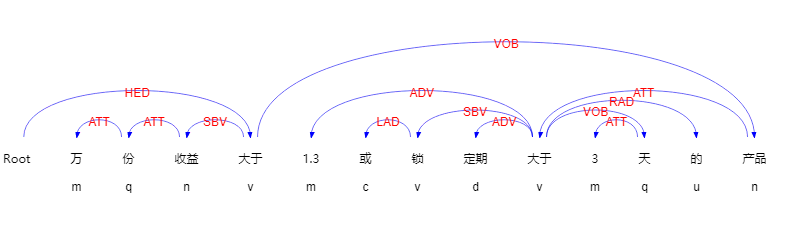
样例后面序号代表对应的工具可识别，序号含义为deep：1；huan：2；ltp：3

1. 万份收益大于1.3或锁定期大于3天的产品（1）
2. 万份收益大于1.3和锁定期大于3天的产品（1）
3. 锁定期大于1.3 收益率大于3天的产品（1）
4. 锁定期大于1.3，收益率大于3天的产品（1，3）
5. 万份收益大于1.3、锁定期大于3天的产品
6. 万份收益大于1.3，锁定期大于3天的产品
7. 万份收益大于1.3 锁定期大于3天的产品
8. 收益率大于5%或理财期限小于30天的产品
9. 万份收益大于1.3、锁定期大于3天、理财期限小于30天的产品
10. 收益率大于5%、起购金额小于5万
11. 收益率大于5%起购金额小于5万（1,2）
12. 收益率大于5%且起购金额小于5万（1,2）
13. 收益率大于5% 起购金额小于5万（1,2）
14. 大于5% 起购金额小于5万（1,2）
15. 收益率大于5%或起购金额小于5万（2）
16. 大于5%、理财期限小于30天（2）
17. 万份收益、起购金额大于1的产品（2）
18. 大于5%、小于30天的理财产品
19. 大于5%小于30天的产品
20. 锁定期和理财期限小于30天

样例具体解析结果如下例子。

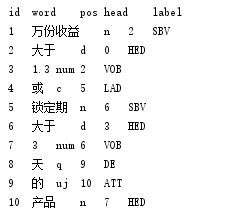
例子：万份收益大于1.3或锁定期大于3天的产品

哈工大语言云分析结果：



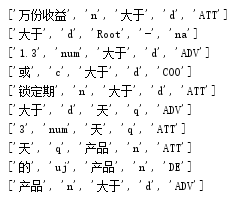
前半部分解析错误，“大于”宾语应为“1.3”

* deepnlp分析结果：



解析正确

* huannlp分析结果：



“或”后面部分解析错误

对20条多关键词查询语句进行验证，三种工具对其解析的正确率如下

|  |  |
| --- | --- |
| 依存句法工具 | 正确率 |
| 哈工大语言云 | 5% |
| deepnlp | 40% |
| huannlp | 35% |

综上，三种工具效果均不理想，主要表现在对关键词敏感，对标点符号不识别，和一些其他原因，导致解析结果不理想。

## 自定义方法

### 判断及解析逻辑

1. 识别是否为多关键词查询

第一种处理方法：

限制：要求多关键词查询不能省略关键词。

如何识别省略关键词，是单关键词查询还是多关键词查询

例：收益率大于5%小于6%的产品

大于5%理财期限小于30天的产品

解析：上面两种情况均含有一个关键词，但一个为单关键词查询，一个为多关键词查询

直接根据关键词的个数判定是否为多关键词查询。

第二种处理方法：

不添加限制

根据att\_name和att\_value判断，若有一个特征的个数大于1，则为多关键词查询

例子：收益率大于5%、理财期限小于30天（两者个数都大于1）。

大于5%、理财期限小于30天（value个数大于1）。

万份收益、起购金额大于1的产品（name的个数大于1）。

但可能出现判断错误的情形。

例如：收益率大于5%小于7%的产品。

一个name，两个value，但此种为单关键词查询，为此，可以增加一个限制，如果只有一个name，且value多个，那么如果value的格式不同，也判定为多关键词查询。

1. 识别布尔逻辑

若有明显表明or的词（如“或”，“或者”，“或是”），则布尔逻辑为or

否则为and（可能的情形，包含明显表明and的词，省略标志词，都默认为and类型）。

1. 将查询语句切成多个单一关键词查询语句

几种情况：

1）有明显的分隔标识，如“，”、“或”、“和”、空格等，直接用分隔标识分隔

1. 没有明显的分隔标识

Name和value都有

Name有省略：根据单关键词

Value有省略：无法继续，提示用户补全信息

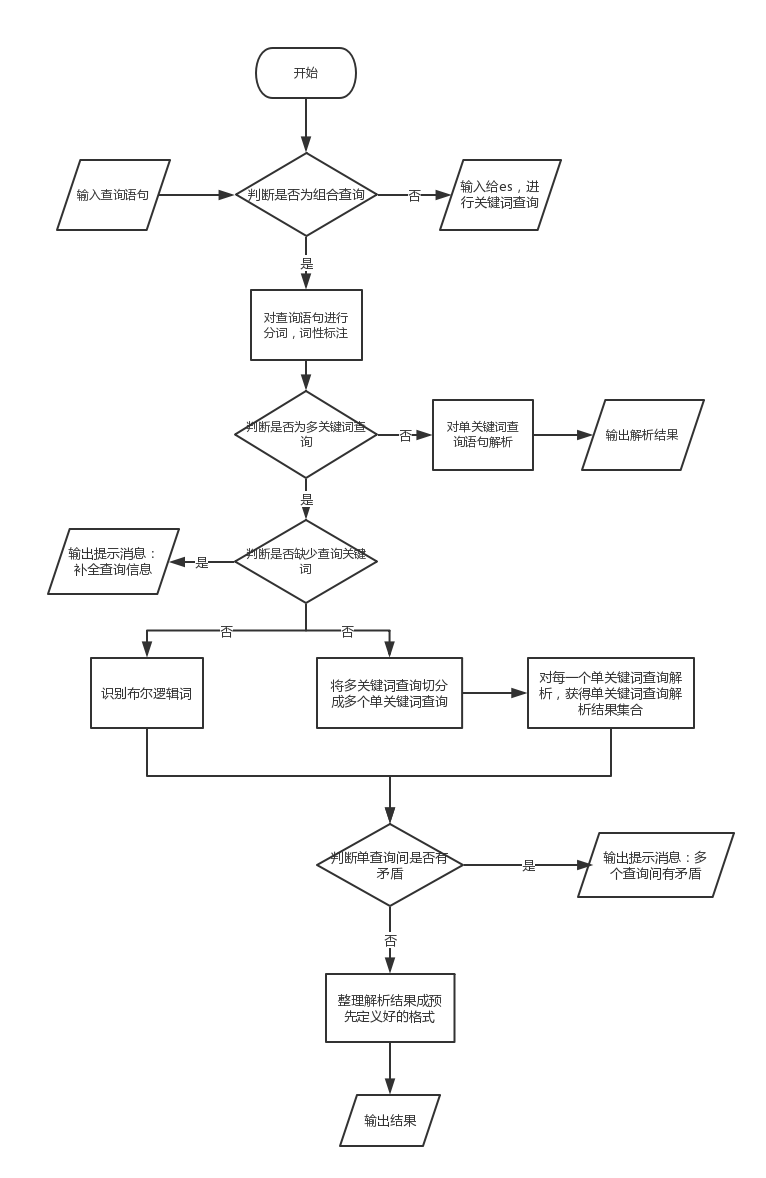
1. 解析切割的单一关键词查询语句

方法见单一关键词组合查询的解析方法，包含各种矛盾的处理及提示

1. 多个单一查询结果产品类型是否矛盾判断

Or：可以出现单一关键词查询结果中产品类型不同

And：不能出现上面or中的情况，若出现提示用户相关信息。



### 实验结果

同样对上述20条样例进行测试，正确率为：80%（红色字体的样例为识别错误的）

## 总结

效率

程序执行时间上自定义更具优势

准确率