课程设计-用户手册

<文件说明>

1. networkhw.py 源代码

2. ruc\_news.txt 原始文本文件

3. \_person.txt 提取出的人物信息文本文件，内容为（人物名，出现频数）

4, \_organization.txt 提取出的机构信息文本文件，内容为（机构名，出现频数）

5. \_relation.txt 用来建立人物关系图的文本文件，内容为（人物名，人物名，边权）

6. \_relation2.txt 用来建立机构关系图的文本文件，内容为（机构名，机构名，边权）

7. stopwords.txt 内容为jieba分词中忽略的词的文本文件

8. userdict.txt 用户自定义词典，用于jieba分词

9. res.csv 记录程序输出

<程序运行说明>

（一）

1. 在命令提示符中打开，进入包含有数据文件和源代码的文件夹。

2. 在命令提示符中键入”python networkhw.py”，程序开始运行。

3. python版本：3.7.1

4. 程序运行时，却确保源代码与数据在同一文件夹下。

（二）

1. 程序运行后，首先选择是从ruc\_news.txt从头开始处理数据，直接从\_person.txt, \_organization.txt, \_relation.txt, \_relation2.txt读取数据。选择前者，则键入1；选择后者，则键入2.

2. 数据加载完毕后，程序自动进行准确性验证，此处将会输出出现频率最高的前10个人物和机构。如果认为准确性可以接受，则在屏幕上打出“准确性验证通过？y/n”后键入“y”，否则键入“n”，程序结束。

3. 准确性验证后，程序开始建立人物图和机构图，并输出与图有关的信息，如节点数、边数、最大连通子图大小等。

4. 图建立完成后，程序给出菜单，用户根据菜单上的选项键入相应的指令。

若键入1，则在“请输入想查询的人物：”后输入人物的名字，如“靳诺”，程序输出与该人物关联边最多的前10个人。

若键入2，则程序输出人物图中pagerank排名前20位。

若键入3，则程序输出中介中心性排名前10位。

若键入4，则程序输出聚集系数排名前10位。

若键入5，则在“请输入选择节点的出现频数最小值：”后输入频数下限，如10,20,50，程序输出由频数不小于频数下限的节点构成的节点对，它们之间共同好友数的统计值。其中，10表示共同好友数大于等于10。

若键入6，则程序输出机构图中pagerank排名前20位。

若键入7，则进入强影响力人物和强影响力机构之间的关联性挖掘程序（详见实验报告4.1.（6））。此时，若之前没有计算过人物图或机构图的pagerank，则会在这一步中进行计算，以得到强影响力人物、机构列表。若再键入1，则程序会输出处理新闻文本的进度（每处理一千行新闻文本报告一次），并最终输出在一篇新闻中，若出现了影响力排在前20为的机构名，则出现影响力排名前20位人物名的概率；键入2，则程序会输出处理新闻文本的进度（每处理一千行新闻文本报告一次），并最终输出在一篇新闻中，若出现了影响力排在前20为的人名，则出现影响力排名前20位机构名的概率。

若键入0，则退出程序。