


第一讲 课程导言与分词


学习使用在线NLPIR分词系统或微词云分词或清华大学分词演示系统（案例演示截图）





微词云
fenci.weiciyun.com


中文分词


搜索问题  Ctrl K

视频教程 

去水印，解锁高级分析 

保存到云端 

其他产品 



中国工程院
研究员
研究室
揭阳市
当选
集团公司
研究
历任
北京
生于
日出
中船
原籍
广东省
毕业
汕尾市
核潜艇
海军
上海交通大学
重工
设计所
总工程师
总体

| 单词 | 全部 | 数量 | 条数 | tf-idf | 操作 |
|--------|-------|----|----|---------------|----|
| 毕业 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 上海交通大学 | 机构团体名 | 1 | 1 | -0.0107510713 | 查看 |
| 历任 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 北京 | 地名 | 1 | 1 | -0.0107510713 | 查看 |
| 海军 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 研究室 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 总工程师 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 中船 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 重工 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 集团公司 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 总体 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 研究 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 设计所 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 研究员 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 名誉 | 名词 | 1 | 1 | -0.0107510713 | 查看 |
| 所长 | 名词 | 1 | 1 | -0.0107510713 | 查看 |

名词 更改词性



核潜艇
院士
所长
名誉
研究员
设计所
总体
集团公司
中船
总工程师
研究室
海军

第一讲 课程导言与分词

安装python (anaconda) (编写输出 "Hello World. Hello '你的姓名' ") ;



The screenshot shows a code editor with a file named `test.py`. The code contains a single line: `print("hello world,魏述桓")`. The editor interface includes a sidebar with icons for Explorer, Search, Source Control, and Run and Debug. The bottom panel shows the output of running the script, displaying the command `C:/python/python.exe e:/souce/Python/test.py` and the output `hello world,魏述桓`.

```
test.py M X
Python > test.py
You, 现在 | 1 author (You)
1 print("hello world,魏述桓") You, 现在 • Uncommitted changes

问题 2 输出 调试控制台 终端 端口 GITLENS

Microsoft Windows [版本 10.0.26200.7462]
(c) Microsoft Corporation。保留所有权利。

E:\souce>C:/python/python.exe e:/souce/Python/test.py
hello world,魏述桓

E:\souce>
```

第一讲 课程导言与分词

课后作业001

子/说/三个/字/：/‘/我爱你/’/如果/非要/给/这份/爱/加上/一个/期限/我/希望/一万年/

- 4.可以开启你的小组项目的第一个小小任务啦！就是对一小段有关“功勋科学家”的文本进行分词处理。请对以上第三段文本进行分词，并评估分词效果（哪些地方分的好，哪些分的不好）。

```
[19]: set_list2 = jieba.cut("黄旭华，1926年3月12日出生于广东省汕尾市，原籍广东省揭阳市。1949年毕业于上海交通大学。历任北京海军核潜艇研究室副总工程师、中船重工集团公司核潜艇
```

```
[22]: print(' '.join(set_list2))
```

黄旭华 ， 1926 年 3 月 12 日 出 生 于 广 东 省 汕 尾 市 ， 原 籍 广 东 省 揭 阳 市 。 1949 年 毕 业 于 上 海 交 通 大 学 。 历 任 北 京 海 军 核 潜 艇 研 究 室 副 总 工 程 师 、 中 船 重 工 集 团 公 司 核 潜 艇 总 体 研 究 设 计 所 研 究 员 、 名 誉 所 长 。 1994 年 当 选 为 中 国 工 程 院 院 士 。

[]:



第一讲 课程导言与分词

课后作业002

jupyter 002-word_cut_科学家文本 Last Checkpoint: 7 months ago



File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel) Anaconda Toolbox

```
[25]: # stopwords = [line.strip() for line in open('stop_words.txt', 'r', encoding='utf-8').readlines()]
```

```
[26]: stopwords = open('stop_words.txt', 'r', encoding='utf-8').read()
stopwords = stopwords.split('\n')
```

```
[27]: stopwords
```

```
[27]: ['的', '了', '是', '啊', '、', '，', '。', '！', '停用']
```

```
[28]: seg_list_huang = jieba.cut('黄旭华，1926年3月12日出生于广东省汕尾市，原籍广东省揭阳市。1949年毕业于上海交通大学。历任北京海军核潜艇研究室副总工程师、中船重工集团公司
```

```
[17]: final = ''
```

```
[18]: for seg in seg_list_huang:
    if seg not in stopwords:
        final += seg + '/'
```

```
[19]: print(final)
```

黄旭华/1926/年/3/月/12/日出/生于/广东省/汕尾市/原籍/广东省/揭阳市/1949/年/毕业/于/上海交通大学/历任/北京/海军/核潜艇/研究室/副/总工程师/中船重工集团公司/核潜艇/总体/研究/设计所/研究员/名誉/所长/1994/年/当选/为/中国工程院院士/

作业的意义：

第一讲 课程导言与分词

课后作业003

Jupyter 003-NER-企业年报-数字技术-安全管理 Last Checkpoint: 7 months ago



File Edit View Run Kernel Settings Help

Not Trusted

Code

JupyterLab Python 3 (ipykernel) Anaconda Toolbox

```
[16]: # 输出所有词汇的词频（按频率降序）
print("\n所有词汇词频统计（前20个）：")
for word, count in word_counts.most_common(20):
    print(f"{word}': {count}次")
```

所有词汇词频统计（前20个）：

```
'，': 13次
'：': 9次
'管理': 5次
'、': 5次
'与': 4次
'“': 3次
'企业': 3次
'”': 3次
'体系': 3次
'智能化': 3次
'经营': 3次
'打造': 3次
'能力': 3次
'供应链': 3次
'建设': 2次
'通过': 2次
'自动化': 2次
'数字化': 2次
'升级': 2次
'生产': 2次
```

第一讲 课程导言与分词

课后作业004

Jupyter 004_使用大语言模型提取科技文献中的实体 Last Checkpoint: 7 months ago



File Edit View Run Kernel Settings Help

Not Trusted

Code

JupyterLab Python 3 (ipykernel) Anaconda Toolbox

```
[13]: # 处理响应
if response.status_code == 200:
    result = response.json()
    try:
        entities = result['choices'][0]['message']['content']
        print("提取到的实体和专业术语:")
        print(entities)
    except KeyError:
        print("无法解析API响应, 原始响应:")
        print(result)
else:
    print(f"请求失败, 状态码: {response.status_code}")
    print(response.text)
```

提取到的实体和专业术语:

```
```json
{
 "理论": ["肿瘤免疫微环境", "T细胞耗竭", "免疫编辑理论"],
 "方法": ["单细胞RNA测序", "细胞亚群聚类", "轨迹分析", "pseudotime推断", "细胞间通讯网络构建"],
 "工具": ["Seurat", "Monocle3", "CellChat"],
 "信号通路": ["PD-1/PD-L1", "TGF-β路径"]
}
```
```

第一讲 课程导言与分词

课后作业004

Jupyter 004_使用大语言模型提取科技文献中的实体 Last Checkpoint: 7 months ago



File Edit View Run Kernel Settings Help

Not Trusted

Code

JupyterLab Python 3 (ipykernel) Anaconda Toolbox

```
[13]: # 处理响应
if response.status_code == 200:
    result = response.json()
    try:
        entities = result['choices'][0]['message']['content']
        print("提取到的实体和专业术语:")
        print(entities)
    except KeyError:
        print("无法解析API响应, 原始响应:")
        print(result)
else:
    print(f"请求失败, 状态码: {response.status_code}")
    print(response.text)
```

提取到的实体和专业术语:

```
```json
{
 "理论": ["肿瘤免疫微环境", "T细胞耗竭", "免疫编辑理论"],
 "方法": ["单细胞RNA测序", "细胞亚群聚类", "轨迹分析", "pseudotime推断", "细胞间通讯网络构建"],
 "工具": ["Seurat", "Monocle3", "CellChat"],
 "信号通路": ["PD-1/PD-L1", "TGF-β路径"]
}
```
```

第一讲 课程导言与分词

阅读压缩文件中（“实体抽取论文-换成PDF”）中的其中一篇论文，并做阅读总结

阅读论文为基于关键词的学术文本聚类集成研究_张颖怡，针对学术文献数量激增、传统分类方法存在局限的问题，本研究提出并验证了一种基于关键词的聚类集成方法，以提升学术文本自动主题划分的性能。研究采用TF-ISF、CSI、ECC和TextRank四种无监督方法抽取关键词，并设计了基于K-means的集成方法（ECKM）和基于增量聚类的集成方法（ECIC）进行实验。结果表明，聚类集成方法能显著提升聚类效果，尤其当学科差异较大或关键词数量较少时优势明显；在各类关键词抽取方法中，TextRank表现最优，且增加关键词数量通常有助于提升性能。因此，在实际学术文本分类中，建议优先采用聚类集成方法，并配合TextRank抽取足量关键词，以获得更稳健、准确的分类结果。未来研究可考虑在中文数据集上验证该方法，并融合引文等多源信息进一步优化性能。

第二讲 词频统计

基于CNKI数据库统计分析2014-2024年（近10年）， “信息资源管理” 或 “网络营销” 或其他你感兴趣的~~主题~~变化趋势。

目录

总体趋势

分布

对比分析

主要主题

次要主题

来源类别

学科

研究层次

文献类型

文献来源

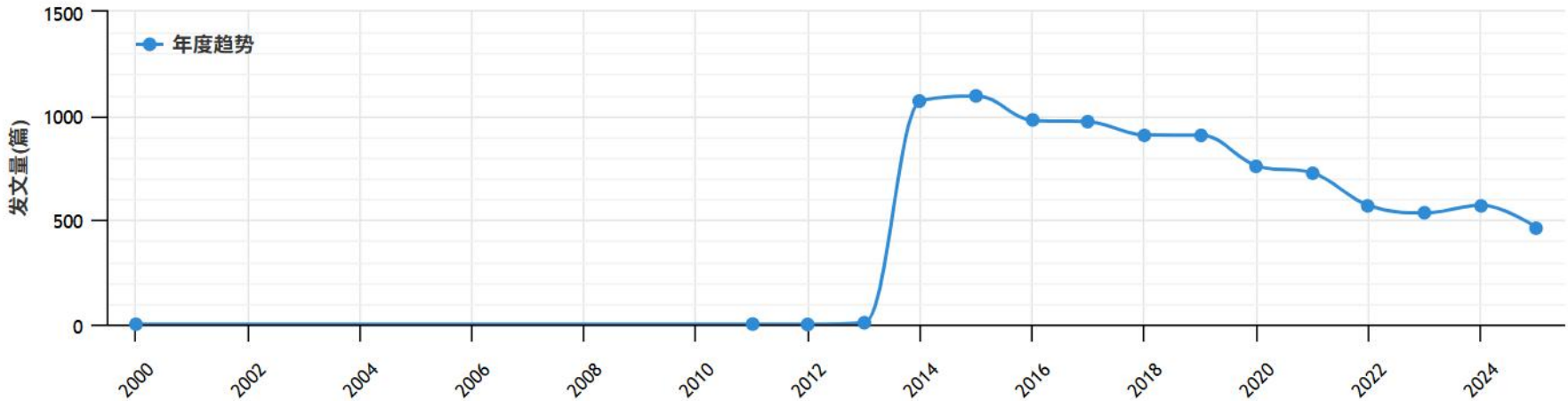
中国作者

机构

基金

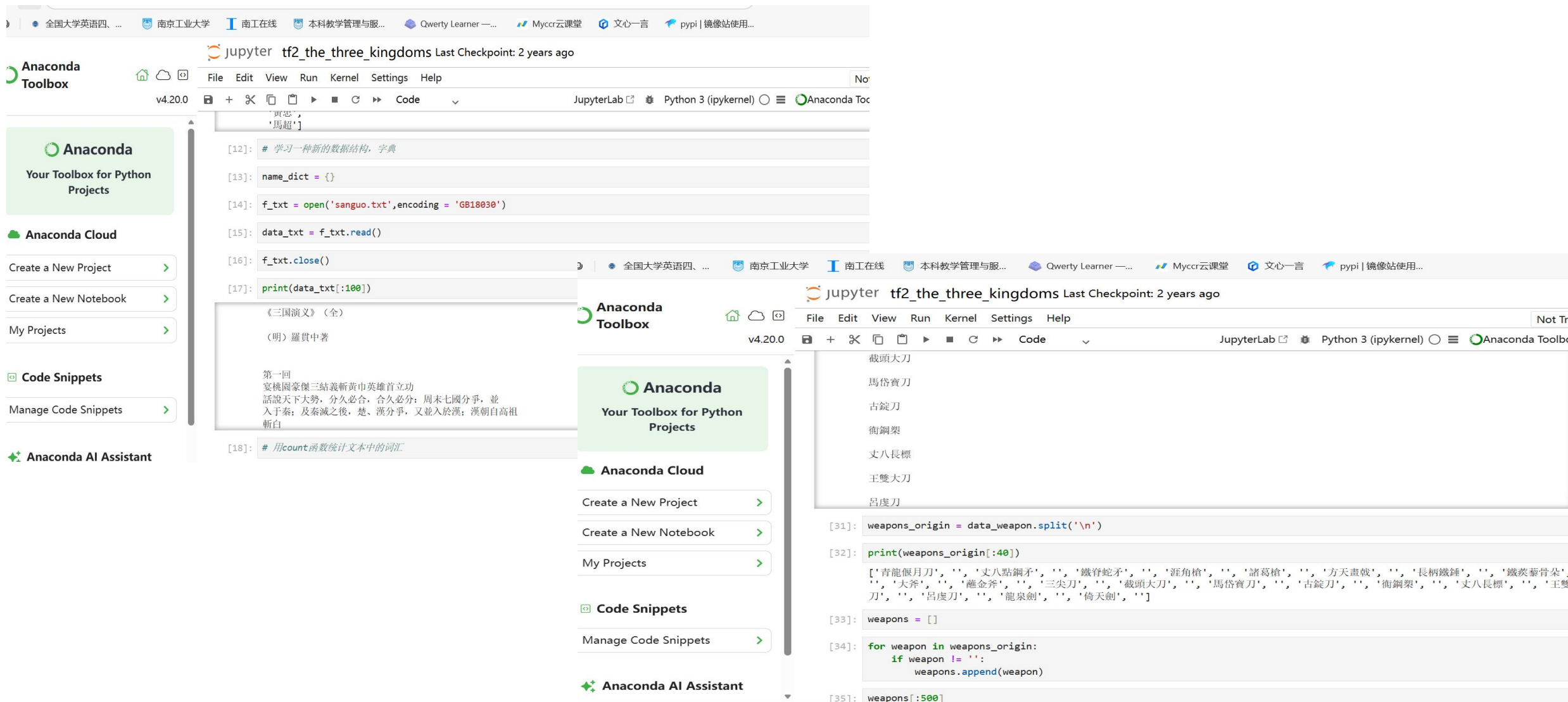
数据来源： 文献总数： 9134 篇； 检索条件： （主题： 信息资源管理）； 检索范围： 总库

总体趋势分析



第二讲 词频统计

完成ppt中的程序运行，包括全文词频统计，指定类型词频统计



The image displays two screenshots of a JupyterLab environment, illustrating the process of reading a text file and performing word frequency analysis.

Left Screenshot: The JupyterLab interface shows a notebook titled "tf2_the_three_kingdoms". The code cells are as follows:

```
[12]: # 学习一种新的数据结构，字典
[13]: name_dict = {}
[14]: f_txt = open('sanguo.txt', encoding = 'GB18030')
[15]: data_txt = f_txt.read()
[16]: f_txt.close()
[17]: print(data_txt[:100])
```

The output of the print statement shows the first 100 characters of the text file, which is the first chapter of "Romance of the Three Kingdoms" (三国演义).

Right Screenshot: The JupyterLab interface shows the same notebook, but with additional code cells for word frequency analysis:

```
[31]: weapons_origin = data_weapon.split('\n')
[32]: print(weapons_origin[:40])
```

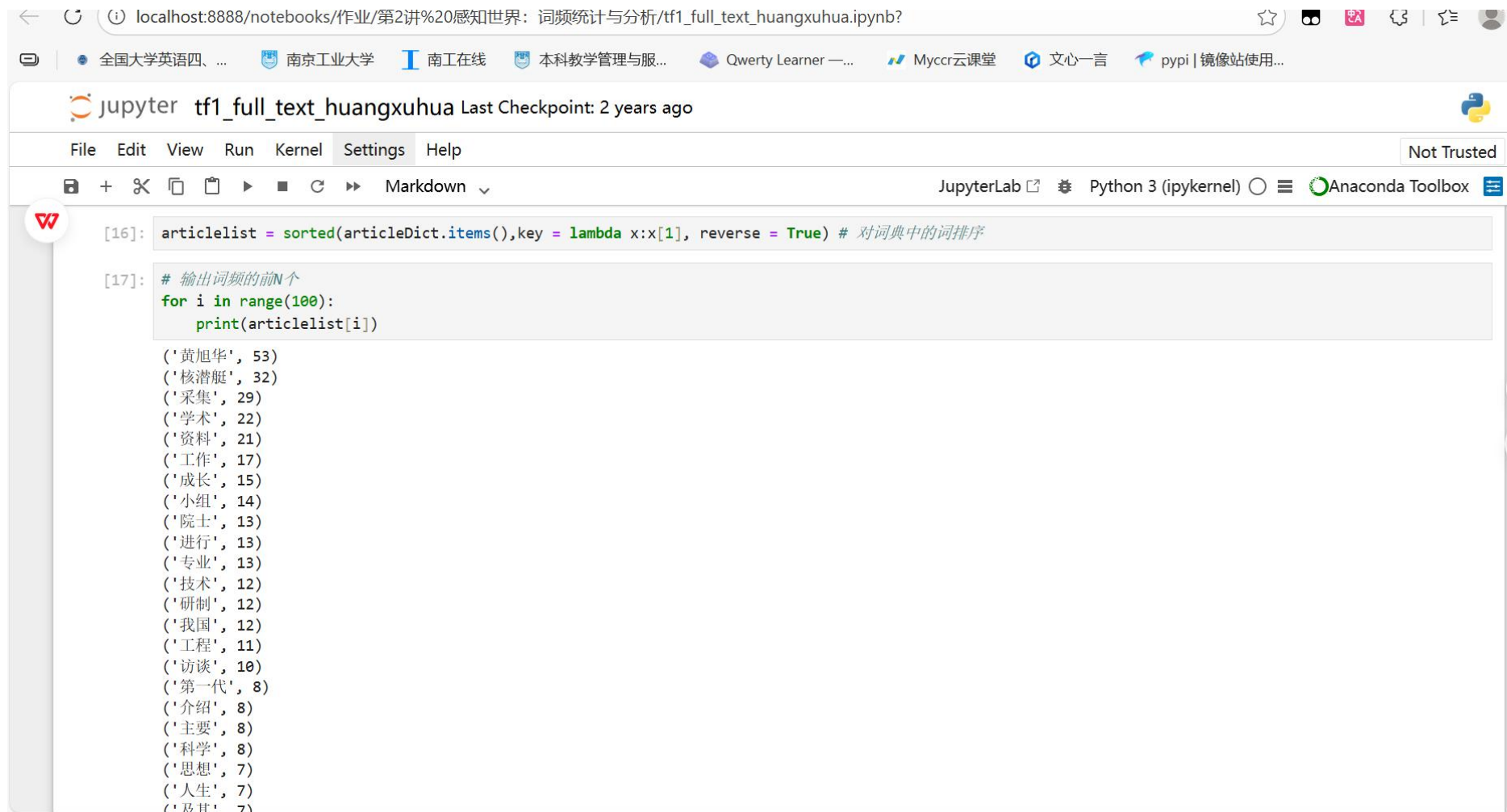
The output of the print statement shows the first 40 lines of the weapons list, which includes various weapons like 截头大刀, 马岱寶刀, 古錠刀, 街鋼槊, 丈八長標, 王雙大刀, and 呂虔刀.

```
[33]: weapons = []
[34]: for weapon in weapons_origin:
[35]:     if weapon != '':
[36]:         weapons.append(weapon)
```

The output of the code shows the list of weapons, which includes: 截头大刀, 马岱寶刀, 古錠刀, 街鋼槊, 丈八長標, 王雙大刀, 呂虔刀, 青龍偃月刀, 丈八點鋼矛, 鐵脊蛇矛, 涯角槍, 諸葛槍, 方天畫戟, 長柄鐵錘, 鐵蒺藜, 大斧, 藤金斧, 三尖刀, 截头大刀, 馬岱寶刀, 古錠刀, 街鋼槊, 丈八長標, 王雙大刀, 呂虔刀, 龍泉劍, 倚天劍.

第二讲 词频统计

链接功勋科学家：把ppt中的文本换成功勋科学家黄旭华院士的传记序言文本



The screenshot shows a JupyterLab environment with a browser window at the top displaying the URL `localhost:8888/notebooks/作业/第2讲%20感知世界：词频统计与分析/tf1_full_text_huangxuhua.ipynb?`. The JupyterLab interface includes a menu bar (File, Edit, View, Run, Kernel, Settings, Help) and a toolbar with icons for file operations and execution. The main area contains two code cells. Cell [16] contains the following Python code:

```
articlelist = sorted(articleDict.items(),key = lambda x:x[1], reverse = True) # 对词典中的词排序
```

. Cell [17] contains the following Python code:

```
# 输出词频的前N个
for i in range(100):
    print(articlelist[i])
```

. The output of cell [17] is a list of words and their frequencies, sorted in descending order of frequency. The output is as follows:

```
('黄旭华', 53)
('核潜艇', 32)
('采集', 29)
('学术', 22)
('资料', 21)
('工作', 17)
('成长', 15)
('小组', 14)
('院士', 13)
('进行', 13)
('专业', 13)
('技术', 12)
('研制', 12)
('我国', 12)
('工程', 11)
('访谈', 10)
('第一代', 8)
('介绍', 8)
('主要', 8)
('科学', 8)
('思想', 7)
('人生', 7)
('及其', 7)
```

第二讲 词频统计

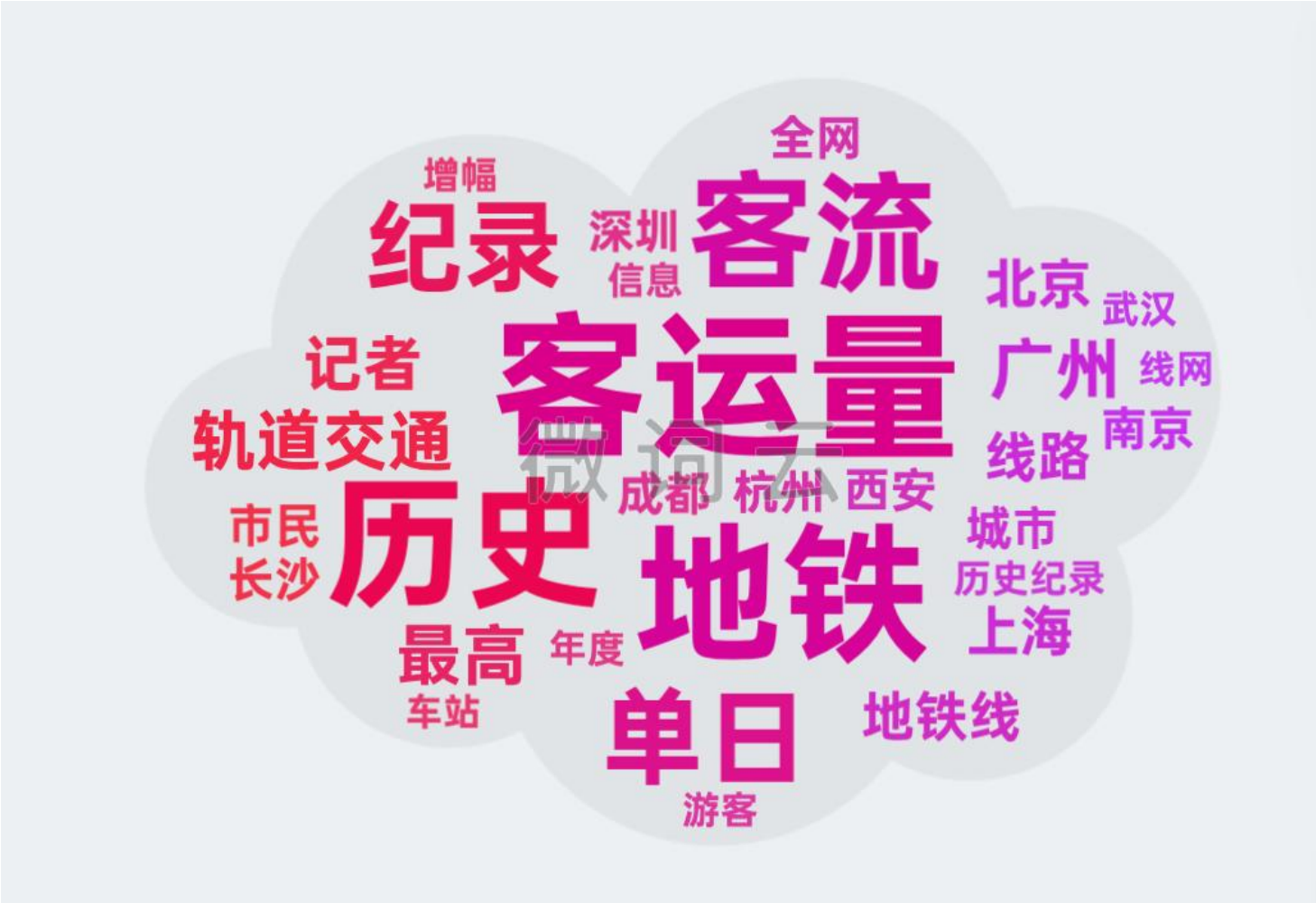
阅读论文 “2018-Wang 等 - Long live the scientists Tracking the scientific” ， 并做阅读总结

- 近年来，随着数字文本资源的普及与计算分析方法的成熟，基于文本量化手段的研究在人文、社会科学乃至计算科学领域展现出日益重要的价值。无论是哲学思想的重构、科学家长期影响力的追踪，还是具体算法在学术研究中的使用评估，都可借助词频统计、提及分析等文本挖掘技术，从海量文献中提取出传统方法难以捕捉的演变规律与深层结构。例如，在福柯研究中，学者通过分析关键词在法文原典中的频次变化，试图还原其思想发展的内在逻辑与阶段转换；在科学计量领域，研究者则利用全球数字化图书与论文数据，以人物名称被提及的频率作为衡量其科学声誉的指标，揭示出科学家影响力的时空分布与学科渗透特征；而在信息科学内部，通过对学术论文全文内容的细粒度分析，可以评估不同算法在特定领域的使用强度、应用场景及随时间的变化趋势，从而为研究者选择工具提供数据支持。
- 这些研究共同体现了数据驱动的学术探索路径：它们不仅关注实体（如概念、人物、算法）的出现频次，更注重其在文本中的位置、语境与演变动态，进而超越主观判断或传统引文分析的局限，形成更为细致、动态且可验证的理解框架。这种跨学科的方法融合，不仅促进了各领域内部的研究深化，也为学术评价、知识演进分析乃至文化趋势洞察提供了新的可能。

第三讲 词云与可视化

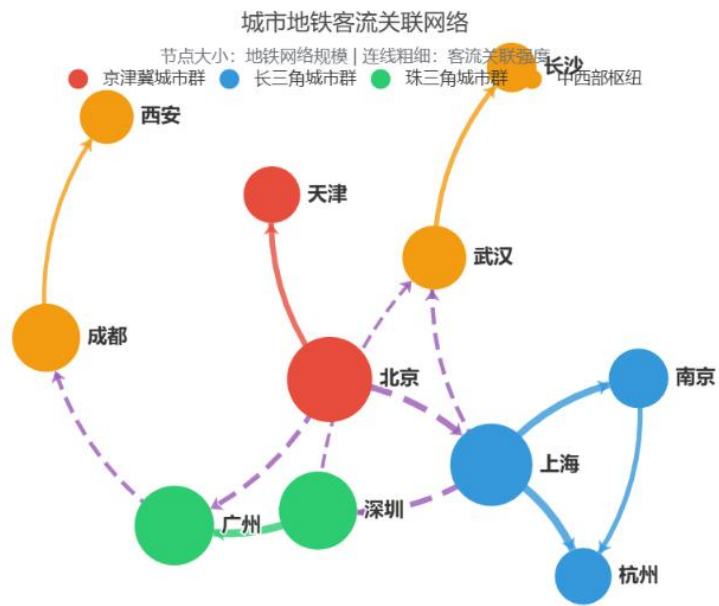
用任意一款词云工具，制作一个好看的词云（内容合理即可），并对词云图有一段话的解释。

这张词云图清晰呈现了近期中国城市轨道交通系统经历的一次全国性客流高峰。核心词汇如“客流”、“历史纪录”、“增幅”和“单日”，直接点明了事件焦点——多个城市的“地铁”客运量在短时间内大幅增长并创下新高。图中并列出现的“深圳”、“北京”、“武汉”、“长沙”、“成都”、“杭州”、“西安”、“南京”等城市名称，表明这是一次广泛存在于主要大城市的普遍现象。与此同时，“全网”、“记者”、“市民”、“游客”等词，则揭示了该事件从媒体广泛报道到公众出行体验的多维度社会关注。整体来看，词云图浓缩了一则关于中国城市化进程中，轨道交通网络运力与公众出行需求持续攀升的生动数据新闻。



第三讲 词云与可视化

使用Echarts，制作3个以上图，其中一个必须是“关系”，图的概念越明确（可解释，而不是自带的模板）越好。



图表解读：

本关系图展示了中国主要城市地铁客流之间的空间关联与网络结构。节点大小代表城市地铁网络规模（线路数量），节点颜色表示所属城市群，连线粗细表示城市间的客流关联强度。可观察到北京、上海、深圳作为三大核心枢纽，与周边城市形成明显的辐射状网络结构。

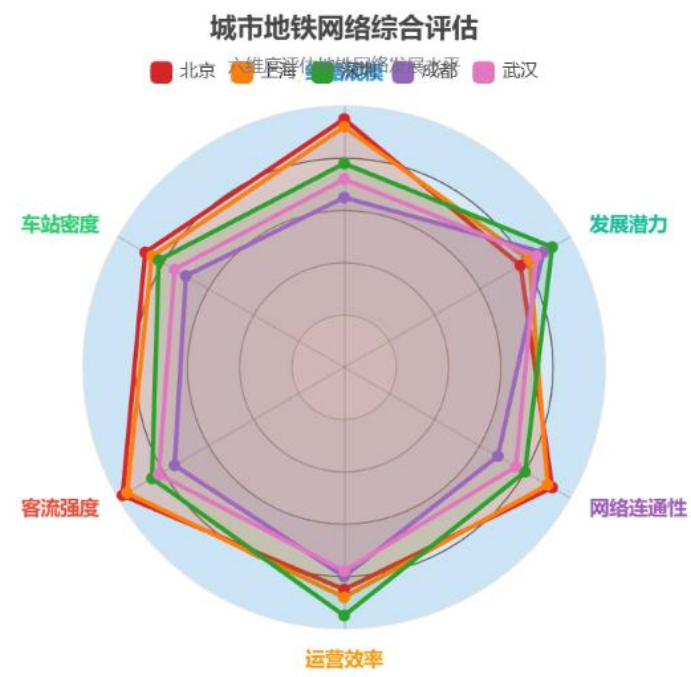


图表解读：

组合图同时展示了各城市地铁单日最高客流量（柱状图）和同比增长率（折线图）。左侧Y轴为客流量（万人次），右侧Y轴为增长率（%）。通过对比可以看出，北京、上海虽然客流量最大，但增长率相对平稳；而成都、武汉等新一线城市则呈现出“高增长、中规模”的特征。

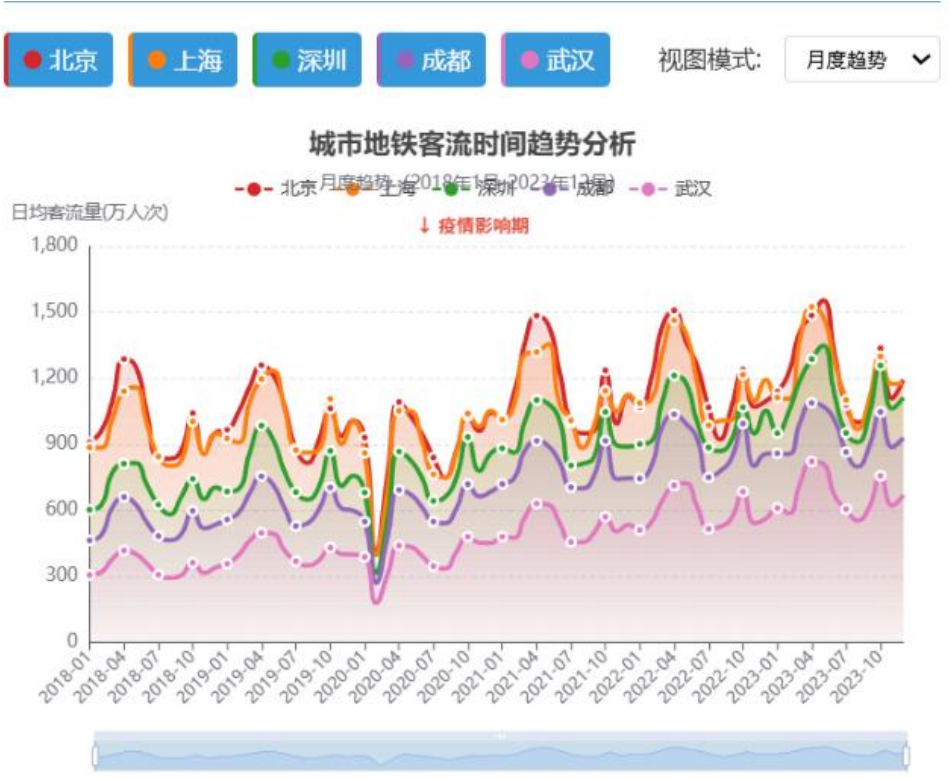
第三讲 词云与可视化

使用Echarts，制作3个以上图，其中一个必须是“关系”，图的概念越明确（可解释，而不是自带的模板）越好。



图表解读：

雷达图从六个维度对城市地铁网络进行综合评估：线路规模、车站密度、客流强度、运营效率、网络连通性和发展潜力。每个维度满分100分，面积越大表示综合发展水平越高。不同形状反映了各城市地铁网络的特色优势，如北京在客流强度上领先，深圳在运营效率上突出。

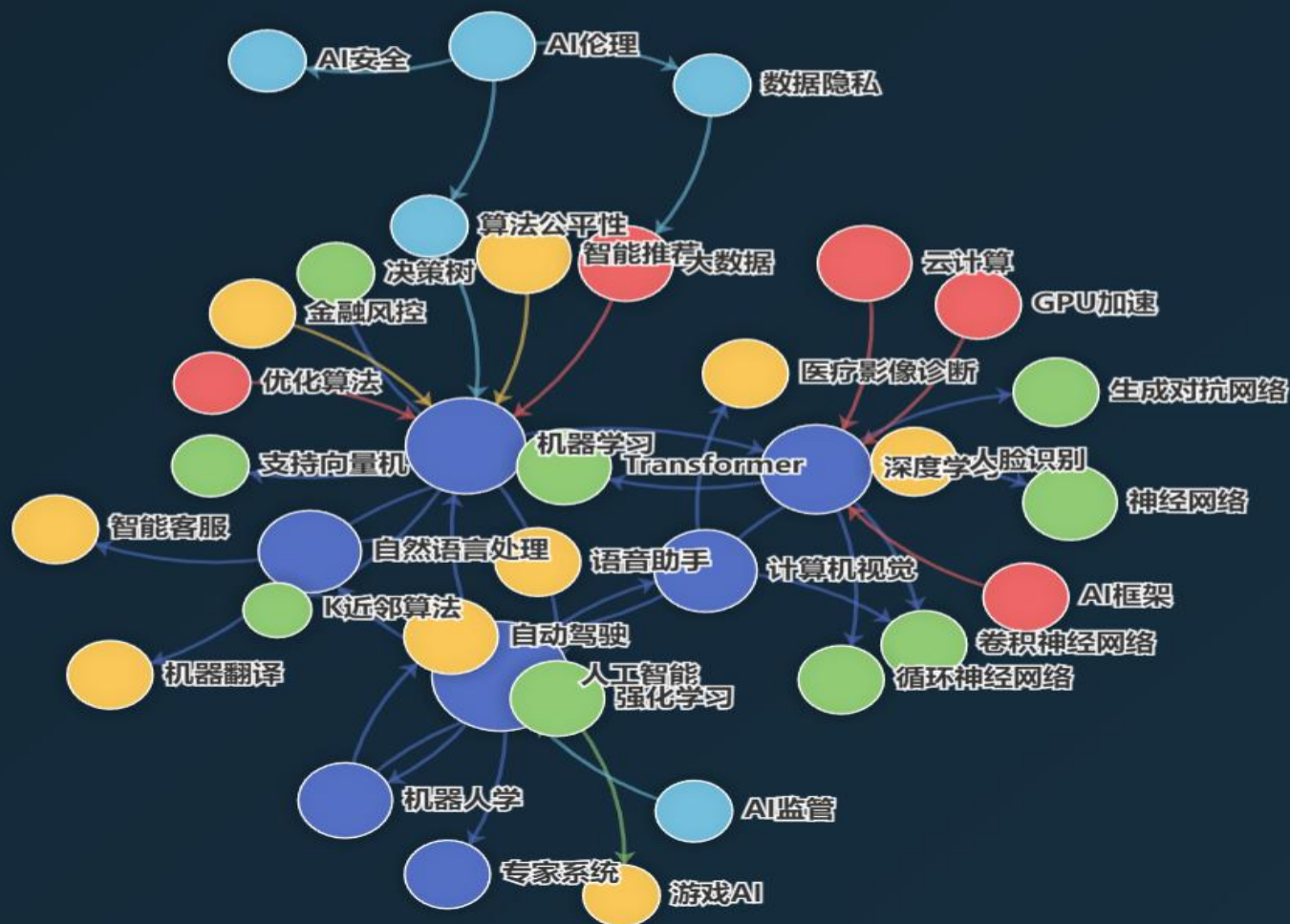


图表解读：

本图展示了主要城市地铁客流随时间的变化趋势。通过切换视图模式可以查看月度、季度或年度数据。图中清晰的季节性波动反映了节假日、天气等因素对地铁客流的影响。2020年初的客流低谷对应了疫情影响，而2021年后的快速恢复则体现了城市交通的韧性。

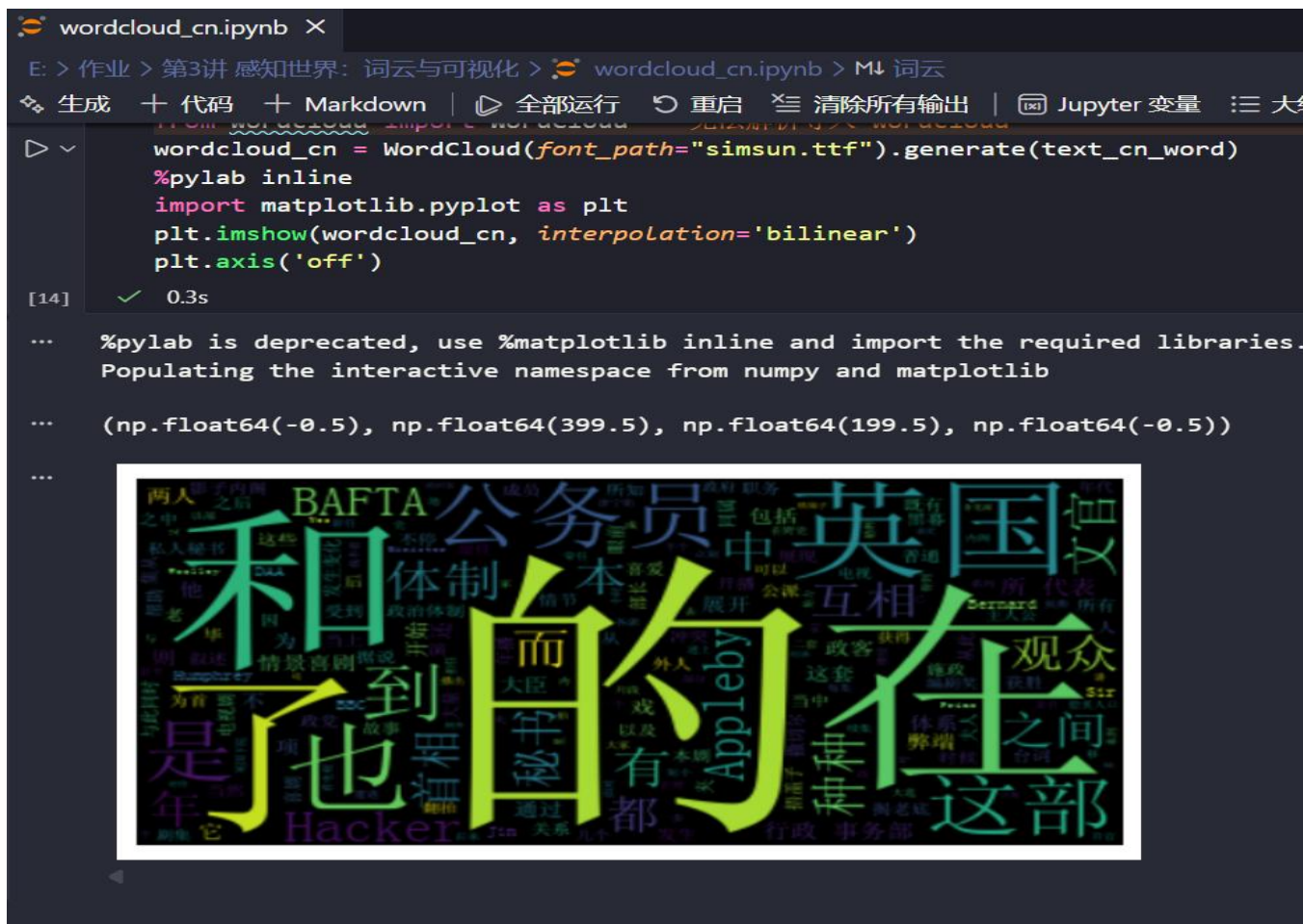
第三讲 词云与可视化

使用Gehpi、VOSViewer、CiteSpace...其中任意一款工具，绘制任意你感兴趣的图谱1-2张



第三讲 词云与可视化

采用给的程序，实现一段科学家文本的词云图绘制，越清晰越好（生成的词云图要单独拿出来）



第四讲 情感分析

使用PPT给的情感分析平台（或其它平台），对文本情感进行分析，并截图；

简体中文 (Chinese Simplified)

要解析的文本或 URL

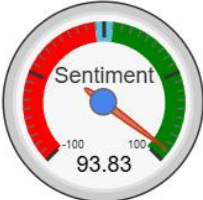
我们继续敞开胸怀拥抱世界。”二〇二〇年，中国始终站在历史正确一边，愿百年变局加速演进，国际形势变乱交北京天安门广场，国之盛典发出正义“中国—中亚精神”熠熠生辉；秋天的

Tonetizer


Sentiment 统计数据

情绪分析

经过处理的文本的总体情绪是。积极的 SWAS: 93.83
SWAS: Saliency Weighted Average Sentiment



| # | 词索引 | Sentiment |
|----|-------|-----------|
| 6 | #国际 🌐 | 积极的 |
| 2 | #人类 🧑 | 积极的 |
| 38 | #高质 🏠 | 积极的 |



第四讲 情感分析

完成sentiment_analysis_1-sentiment_analysis_4，4份代码。

Home

sentiment analysis 1 chuji

+

localhost:8888/notebooks/第4讲%20感知世界：情感分析/第4讲%20感知世界：情感分析/sentiment_analysis_1_chuji.ipynb

jupyter sentiment_analysis_1_chuji Last Checkpoint: 9 days ago

File Edit View Run Kernel Settings Help

Code

```
[17]: textbae_2 = SnowNLP(textbae_2)

[17]: for sentence in textbae_2.sentences:
      print(sentence)
      总结：这是我用过最不再喜欢的一款手机
      两千多元的手机这样
      最垃圾手机
[18]: textbae_2.sentiments
```

Home

sentiment analysis 2 timeline

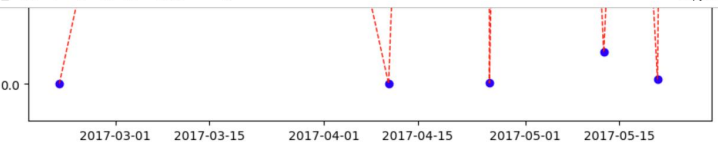
+

localhost:8888/notebooks/第4讲%20感知世界：情感分析/第4讲%20感知世界：情感分析/sentiment_analysis_2_timeline.ipynb

jupyter sentiment_analysis_2_timeline Last Checkpoint: 2 years ago

File Edit View Run Kernel Settings Help

Code



```
[16]: plt.savefig('timeline.png') # 看不到？改一改？
<Figure size 640x480 with 0 Axes>
```

在图中，我们发现许多正面评价情感分析数值极端的高。同时，我们也清晰地发现了那几个数值极低的点。对应评论的情感分析数值接近于0。这几条评论，被Python判定为基本上没有正面情感了。

从时间上看，最近一段时间，几乎每隔几天就会出现一次比较严重的负面评价。

作为经理，你可能如坐针毡。希望尽快了解发生了什么事儿。你不用在数据框或者Excel文件里面一条条翻找情感数值最低的评论。Python数据框Pandas为你提供了非常好的排序功能。假设你希望找到所有评论里情感分析数值最低的那条，可以这样执行：

```
[17]: df.sort_values(['sentiments'])[1:]
```

| | comments | date | sentiments |
|----|--|---------------------|--------------|
| 24 | 这次是在情人节当天过去的，以前从来没在情人节正日子出来过，不是因为没男朋友，而是感觉哪哪儿... | 2017-02-20 16:00:00 | 6.334066e-08 |

情感分析结果数值几乎就是0啊！不过这里数据框显示评论信息不完全。我们需要将评论整体打印出来。

```
[18]: print(df.sort_values(['sentiments']).iloc[0].comments)
```

这次是在情人节当天过去的，以前从来没在情人节正日子出来过，不是因为没男朋友，而是感觉哪哪儿人多，所以特意错开，这次实在是饿A餐厅了，所以赶在正日子也出来了，从下午四点多时候我看排号就排到一百多了，我从家开车过去堵堵的一个小时，我一看提前两个小时就在网上先排着号了，差不多我们是六点半到的，到的时候我看号码前面还有才三十多号，我想着肯定没问题了，等一会就能吃上的，没想到悲剧了，就从我到那坐到等位区开始，大约是十分二十分一叫号，中途多次我都想走了，哈哈，哎，等到最后早上九点才吃上的，服务员感觉也没以前清闲时周到了，不过这肯定的，一人负责好几桌，今天节日这么多人，肯定是很累的，所以大多也都是我自己跑腿，没让服务员给弄太多，就奸滑让服务员下的，然后环境来说感觉卫生方面是不错，就是有些太吵了，味道还是一如既往的那个味道，不过A餐厅最人性化的就是看我们等了两个多小时，上来送了我们一张打折卡，而且当次就可以使用，这点感觉还是不错的，不愧是A餐厅，就是比一般的要人性化，不过这次就是选错日子了，以后还是得提前预约，要不就别赶节日去，太火爆了！

分析

- 通过阅读，你可以发现这位顾客确实有了一次比较糟糕的体验——等候的时间太长了，以至于使用了“悲剧”一词；另外还提及服务不够周到，以及环境吵闹等因素。正是这些词汇的出现，使得分析结果数值非常低。

第六讲 知识图谱理念

LinkedIn知识图谱：当知识图谱遇见大语言模型的进化

- 01 最新进展：安全态势平台（SPP）—— AI驱动的安全防御新范式
- 核心理念演进：从静态的知识网络进化为动态、智能、可交互的主动防御系统。
- 技术深度融合：将知识图谱（存储海量资产与关系事实）与生成式AI（自然语言交互与智能决策）深度融合，构建了安全态势平台（SPP）。
- 核心业务价值：旨在为超过十亿用户和其背后的庞大数字资产构建智能防御体系，应对日益增长的安全威胁。

第六讲 知识图谱理念

LinkedIn知识图谱：当知识图谱遇见大语言模型的进化

- 02 生态构建：以知识图谱为“大脑”，以LLM为“交互界面”
- 坚实的基础层：一个集成了20多个安全数据源的集中化安全知识图谱，收录了从物理设备到云资源的全部数字资产及其相互关系，为风险分析提供了唯一可信源。
- 创新的智能层 (SPP AI)：在知识图谱之上构建AI层，用户可以使用自然语言直接提问（例如，“我们是否受到某漏洞影响？”），系统自动将其转化为图谱查询并生成答案。
- 统一的平台层：提供直观的UI、API及GraphQL沙盒，但AI交互极大降低了使用门槛，使安全分析效率发生质变。

第六讲 知识图谱理念

LinkedIn知识图谱：当知识图谱遇见大语言模型的进化

- 03 关键技术：实现“对话式”安全管理的四大支柱
- 多查询生成：利用LLM将用户的模糊需求，分解并映射为知识图谱中精确的实体、关系和查询语句（如Cypher查询）。
- 上下文生成与检索增强：通过合成数据生成与向量索引，动态构建丰富的提示上下文，有效应对复杂、罕见的查询场景，减少AI“幻觉”。
- 严格的评估框架：采用包含“种子数据验证”、“人工复核”在内的多层评估体系，确保AI生成结果的准确性（据报道，其准确率从早期模型的40%-50%提升至85%-90%）。
- 敏捷的架构设计：采用可插拔设计，能够快速集成新的数据源和适配最新的大语言模型（如从GPT-3迭代至GPT-4），保持技术前瞻性。

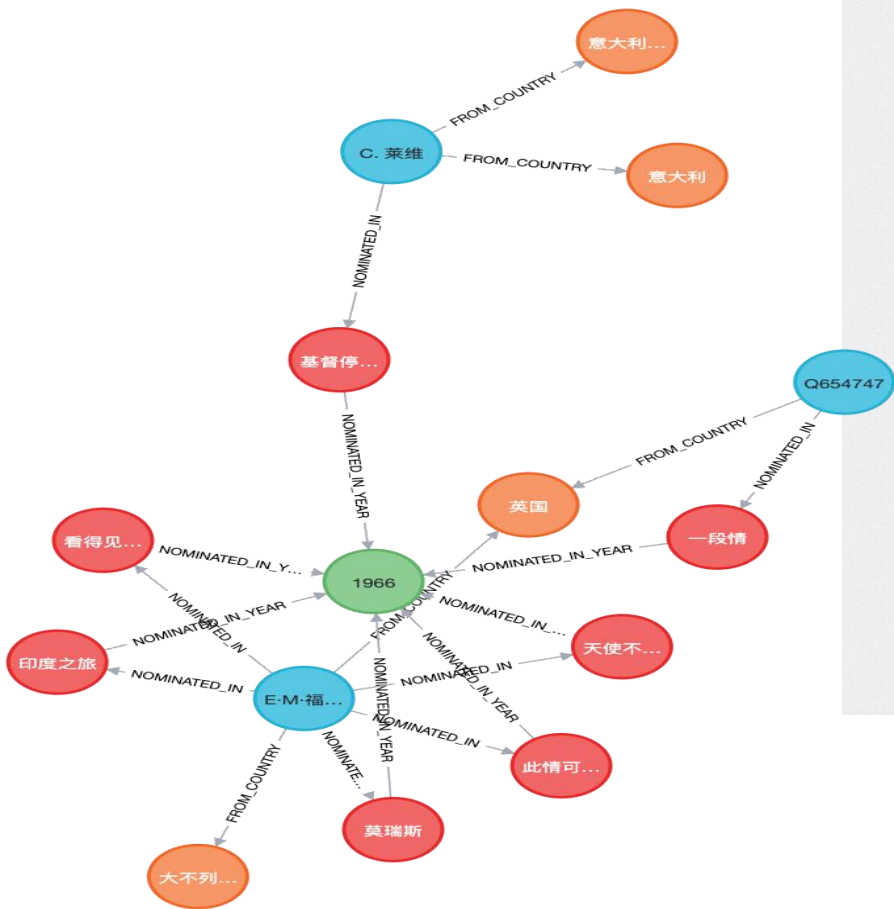
第六讲 知识图谱理念

LinkedIn知识图谱：当知识图谱遇见大语言模型的进化

- 04 特色与评价：从“数据关联”到“智能认知”的跨越
- 特色一：问题导向，价值驱动：不以构建庞大图谱本身为目标，而是直击“安全漏洞响应效率低下”这一具体业务痛点，将知识图谱作为解决方案的核心基础设施。
- 特色二：交互革命，降低门槛：通过引入LLM，彻底改变了人与知识系统的交互方式，使复杂的安全数据查询变得如同日常对话，极大释放了专业知识的潜力。
- 特色三：前瞻性的架构设计：其系统设计考虑了技术的快速迭代，为持续集成更先进的AI模型留出空间，展现了良好的技术预见性和工程韧性。
- 评价：LinkedIn的知识图谱生态代表了当前该领域一个明确的前沿方向——深度智能化和强业务融合。它成功地将知识图谱从后台的“数据水库”升级为前台的“智能参谋”，其价值已超越传统的信息关联与检索，在于赋能非技术专家进行深度数据洞察与决策。这不仅是技术的胜利，更是产品思维和业务理解的胜利，为知识图谱在复杂企业环境中的应用提供了极具参考价值的范式。

第六讲 知识图谱理念

使用PPT中知识图谱链接平台，检索、截图（大词林等，可用的）



诺奖文学提名知识图谱

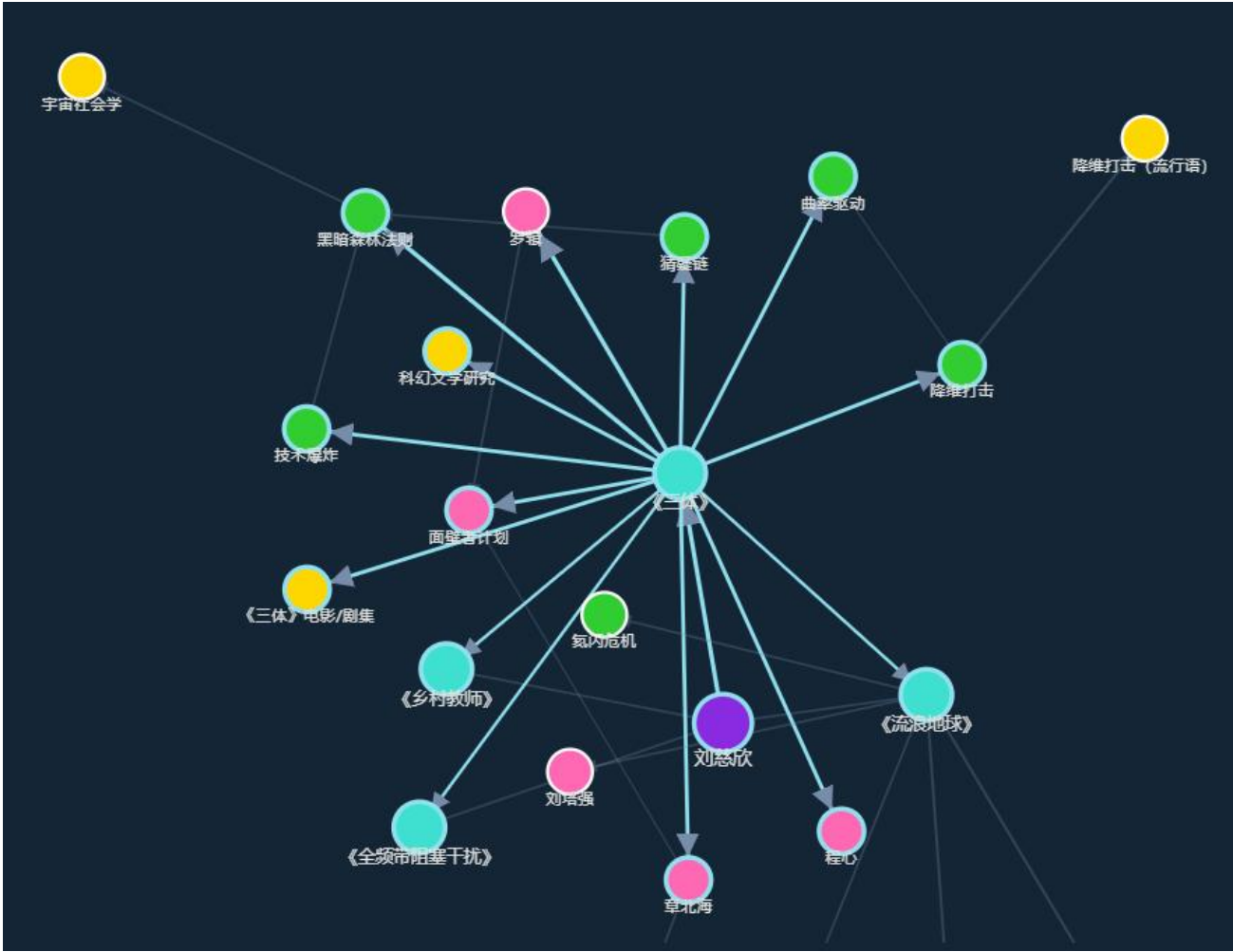
数据与资源

- 诺奖文学提名知识图谱.zip [下载量: 90]
- export.csv [下载量: 71]
- graph.png [下载量: 17]
- nobel.csv [下载量: 39]
- records.json [下载量: 35]
- Readme.md [下载量: 20]
- 2013 诺奖文学提名知识图谱构建课程报告.pdf [下载量: 59]

其他信息

第六讲 知识图谱理念

使用白板建模绘制一个你感兴趣的“知识图谱”，可以是人物关系，也可以是事物关系，或者概念之间的关系等等，并解释你绘制的图谱；



图谱的架构分为四个互联的主干网络：

作品网络：以刘慈欣、郝景芳等作者为核心，连接其创作的《三体》《流浪地球》《北京折叠》等作品。

人物与组织网络：展示如章北海、罗辑等角色，及其隶属的“面壁者计划”等组织。

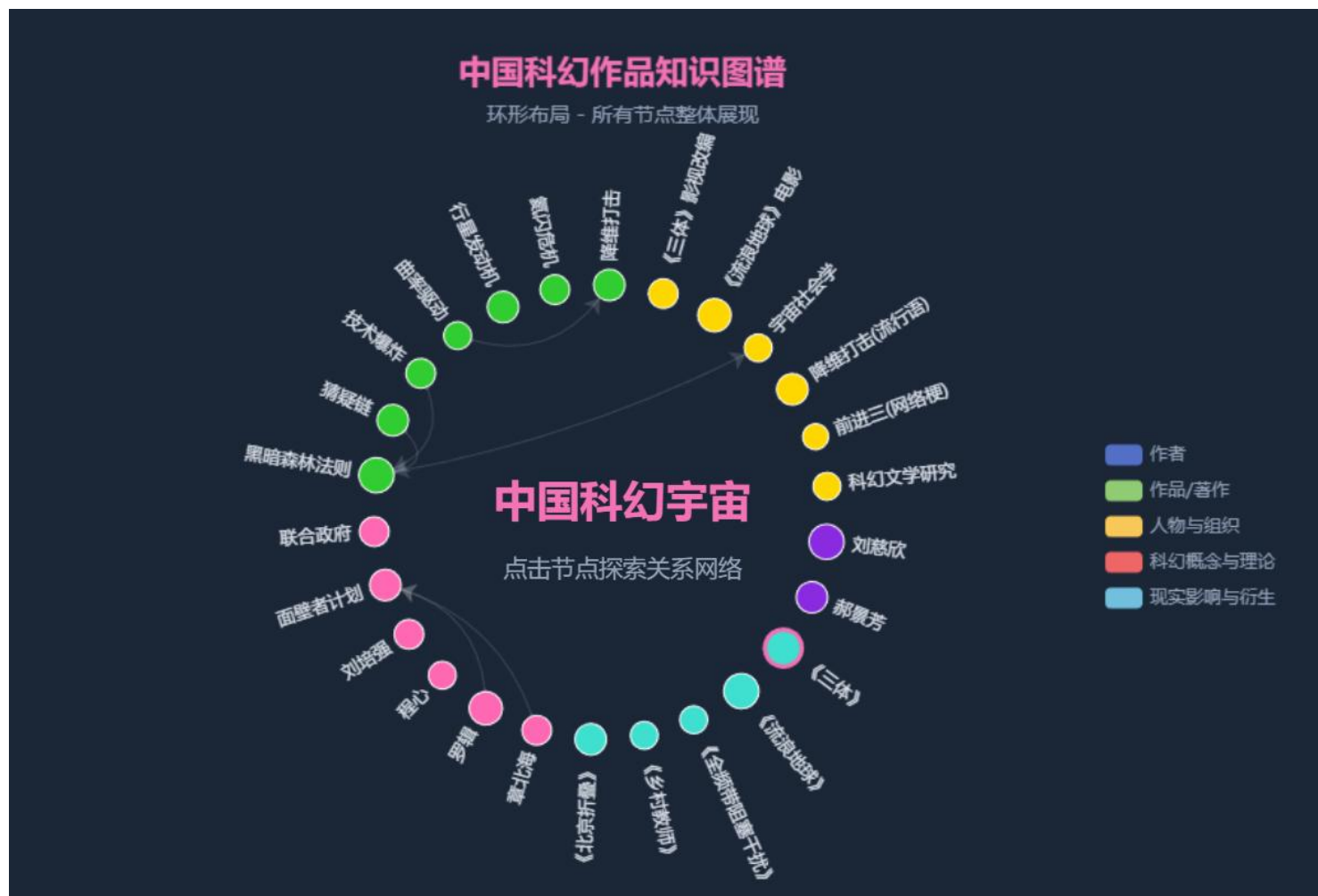
科幻概念网络：呈现“黑暗森林法则”、“猜疑链”、“行星发动机”等核心科幻设定，并揭示概念间的推导关系（如“猜疑链”与“技术爆炸”共同推导出“黑暗森林法则”）。

现实影响网络：追踪这些科幻概念如何溢出文本，演变为“降维打击”等流行语，或启发“宇宙社会学”等学术探讨，并催生影视改编。

其核心价值在于直观揭示了中国科幻的“思想骨架”与虚构与现实的双向滋养过程。通过交互，用户可以探索例如“刘慈欣→《三体》→黑暗森林法则→宇宙社会学”这样的思想传承路径，从而系统性理解中国科幻的叙事体系与思想脉络。

第六讲 知识图谱理念

使用echarts中的关系图，绘制作业2) 中的“知识图谱”



第六讲 知识图谱理念

使用Neo4j（可在线版本），编程绘制一款（简单）知识图谱（内容不限）

