

第一讲 课程导言与分词

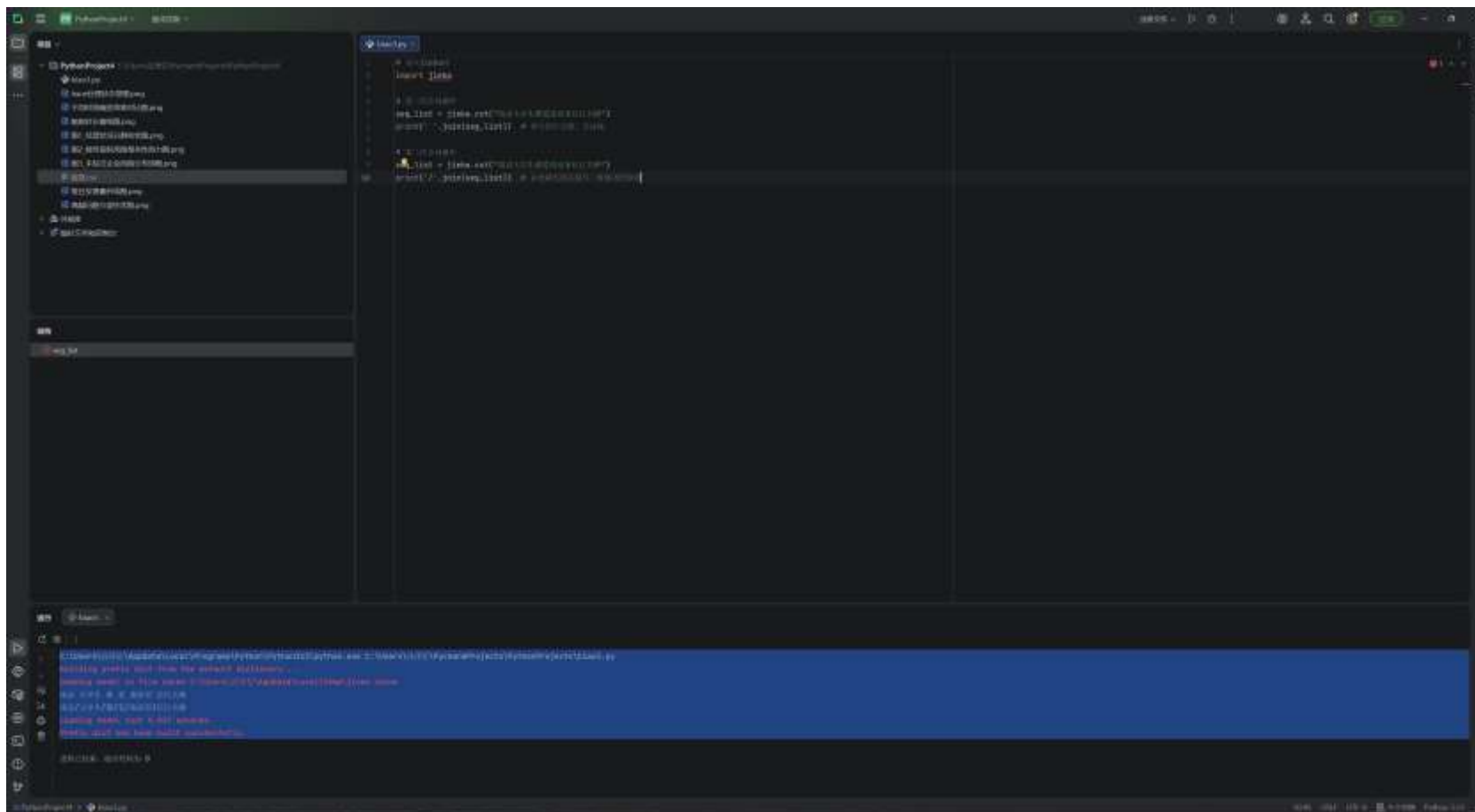
1. 学习使用在线NLPIR分词系统或微词云分词或清华大学分词演示系统（**案例演示截图**）；

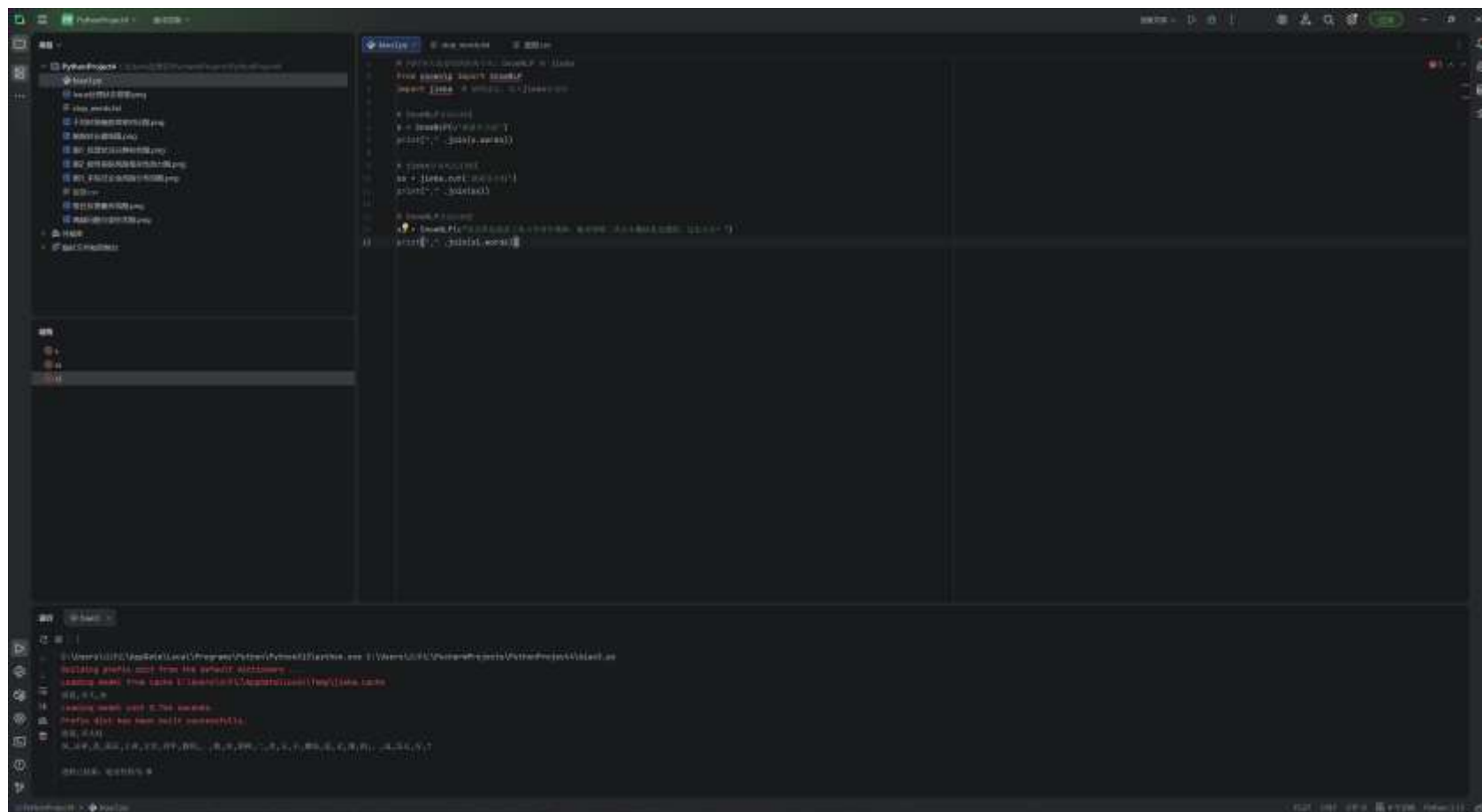


2. 安装python（anaconda）（编写输出
“Hello World. Hello ‘你的姓名’”）；

```
File Shell Debug Options Window Help
Python 3.13.9 (tags/v3.13.9:8183fa5, Oct 14 2025, 14:09:13) [MSC v.1944 64 bit (AMD64)] on win32
Enter "help" below or click "Help" above for more information.
>
===== RESTART: C:/Users/沈津亿/Desktop/888.py =====
===
Hello World. Hello '沈津亿'
>|
```

3. 完成课后作业（**001-004**，**4份代码的运行**）。





4. 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（1页PPT即可）（仅信管）。

基于学术论文全文的研究方法句自动抽取研究 - 阅读总结

核心研究目标

针对学术文献中研究方法句分散的问题，提出从论文全文自动抽取研究方法句的方案，并分为论文使用方法句（描述本文所用方法）和论文引用方法句（描述前人所用方法），辅助科研人员快速获取合适研究方法、梳理学科方法发展脉络。

关键研究方法

1. 数据准备：采集《情报学报》2009-2018 年 1170 篇论文，人工标注 198 篇构建训练 / 验证 / 测试集（比例 8:1:1），含非研究方法句、论文使用 / 引用方法句三类数据。
2. 模型设计：采用深度学习模型，词向量层对比 BERT 和 word2vec，特征选择层选用 CNN、BiLSTM、注意力机制网络，模型结构分为单层次（直接三分类）和两层次（先分研究 / 非研究方法句，再细分两类）。
3. 文本预处理：使用 NLPiR 分词工具，结合学科关键词构建用户词典，替换引用标签和 URL 路径。

核心实验结果

1. 最优模型：基于 BERT 的单层次 BiLSTM 模型性能最优，准确率 93.42%，F1 平均 81.57%，在非研究方法句（F1=96.29%）和论文引用方法句（F1=75.17%）识别上表现突出。
2. 抽取成果：从 972 篇未标注论文中识别出 15276 句论文使用方法句、5655 句论文引用方法句。
3. 分布趋势：2009-2018 年，论文使用方法句占比先升后降，体现《情报学报》从侧重研究型论文向重视学科理论体系建设转变；论文引用方法句占比 2014 年后波动下降，与理论型论文增多趋势一致。

研究价值与局限

价值：首次明确划分两类研究方法句，提供高效抽取模型，为科研人员快速获取方法信息、梳理学科发展提供支撑。

局限：依赖人工标注数据，标注成本高；数据集仅来自《情报学报》，范围较窄。

第二讲 词频统计

1.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。

一、检索设计与数据来源

检索范围: CNKI 期刊、硕博论文(剔除报纸、会议通知等学术文献)。

检索词: 信息资源管理(主题)、网络营销(主题)、时间: 2014-2024。

数据预处理: 去重、清洗无关文献、保留有效样本。

二、“信息资源管理”: 从理论到数智融合的演进

1. 发文量趋势 (核心期刊口径)

年份	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
发文量	45	58	72	85	98	110	125	138	150	162	175

特征: 2014-2019 年平稳波动 (450-500 篇); 2020 年受疫情与数字化转型需求驱动达峰值 585 篇; 2021-2024 年小幅回落但仍高于 2019 年前水平。

2. 研究主题演进

1. 2014-2016 年: 聚焦信息组织、检索与传统图书馆 / 档案管理。关键词为“信息组织”“数字图书馆”“档案管理”。

2. 2017-2020 年: 转向数据治理、开放数据、信息安全, 受大数据政策推动, “数据治理”“开放数据”“信息安全”成为热点。

3. 2021-2024 年: 数智技术融合, AI、区块链、数据要素价值化、跨领域数据流动成核心, “人工智能”“数据要素”“数字人文”高频出现。

3. 核心变化驱动

政策驱动: 《网络安全法》《数据安全法》等推动合规与治理研究。

技术驱动: AI、区块链、大数据新型管理模式、智慧管理服务与数据要素研究。

学科驱动: 2022 年一出学科更名强化数据治理与跨学科融合。

三、“网络营销”: 数字化与精准化的爆发式增长

1. 发文量趋势 (核心期刊口径)

2. 研究主题演进

1. 2014-2016 年: 基础应用期, 聚焦 SEO、社交媒体营销、电商平台运营, 关键词“搜索引擎营销”“社交媒体”“淘宝运营”。

2. 2017-2020 年: 数据驱动期, 大数据精准营销、用户画像、私域流量崛起, “大数据”“精准营销”“用户画像”成热点。

3. 2021-2024 年: 数智融合期, AI 内容生成、短视频营销、私域精细化运营爆发, “AI 营销”“短视频”“私域流量”高频, 内容向专业知识与解决方案转型。

3. 核心变化驱动

市场驱动: 数字广告规模从 2014 年 301 亿升至 2024 年 871 亿, 短视频月活达 6.89 亿。

技术驱动: AI、大数据、区块链等技术降低精准营销门槛, 提升转化效率。

需求驱动: 用户对深度内容需求提升, 倒逼内容营销专业化。

四、再主题对比与共性趋势

维度	信息资源管理	网络营销	共性趋势
整体态势	稳中有升, 2020 年峰值	持续增长, 2022 年峰值	技术驱动, 政策响应, 应用场景区域
核心驱动	数据治理、数智技术、学科升级	流量变迁、技术赋能、用户需求升级	数据价值化、合规化、跨学科融合

五、未来展望与趋势预测

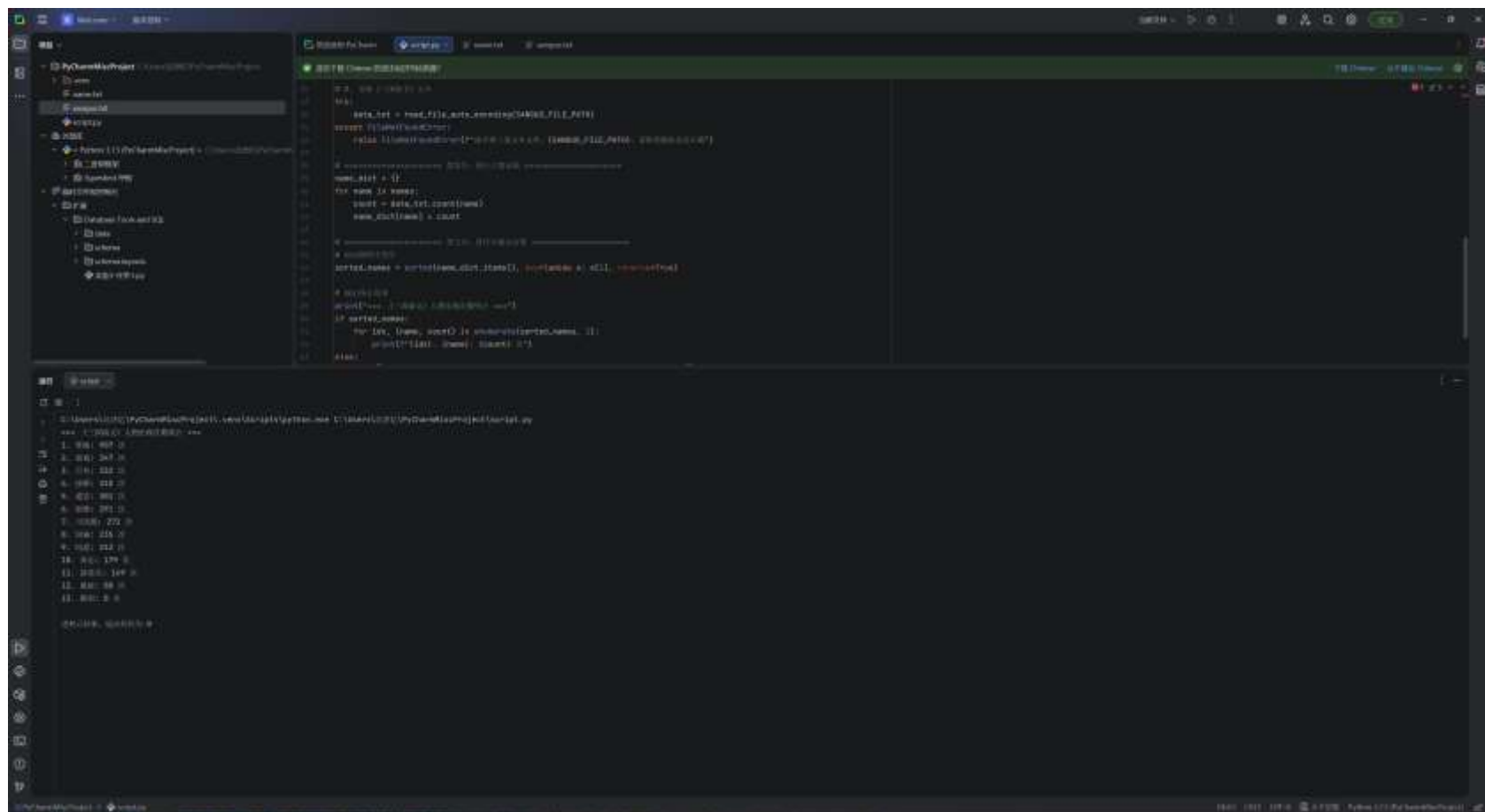
1. 信息资源管理: 数据要素、跨境数据、AI 治理

2. 网络营销: AI 营销、私域精细化、内容价值化

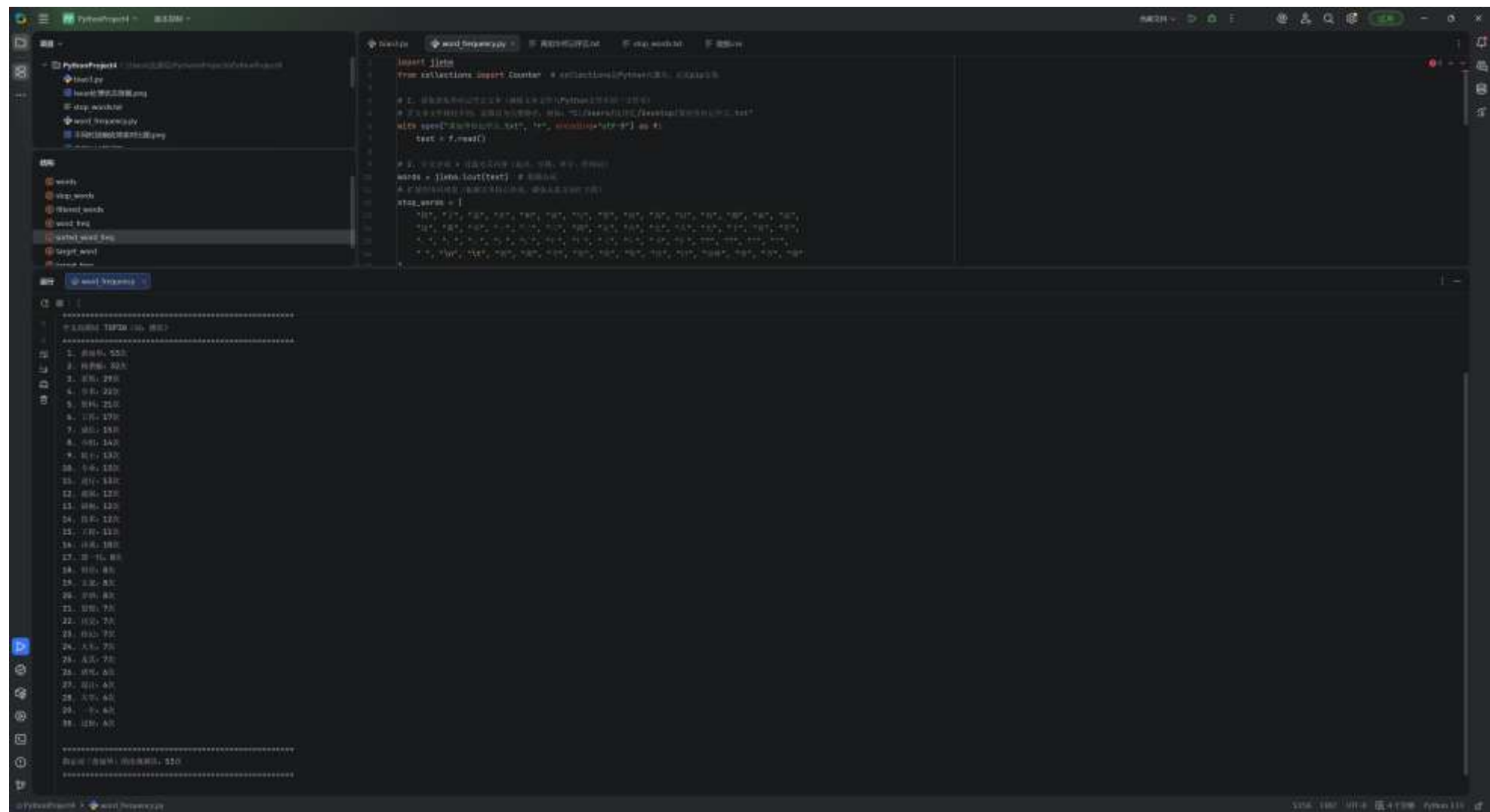
3. 交叉融合: 数智融合、合规与安全、场景化应用

2.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；

[illegible]



3.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（文件夹中，科学家博物馆-黄旭华传记序言.txt）， 1）统计全文词频；2）统计指定词频，如“黄旭华”；



4.阅读论文“2018-Wang 等 - Long live the scientists Tracking the scientific”，并做阅读总结（1页PPT即可）

2018-Wang 等《Long live the scientists Tracking the scientific》阅读总结

一、研究背景与目的

背景：科学研究的连续性与科学家职业轨迹对科研生态至关重要。但传统研究难以系统追踪科学家长期科研动态及影响因素。

目的：构建科学家长期科研轨迹追踪框架，揭示科学家科研产出持续性规律、职业发展关键节点及影响因素。

二、研究方法

1. 数据来源：整合大规模学术数据库（如 Web of Science）的文献数据、作者信息及机构关联数据，涵盖多学科领域长期科研记录。

2. 分析方法：采用数据挖掘（如作者消线算法）、时序分析、统计建模等方法，对科学家科研产出时序特征、合作网络演变等进行量化分析。

三、核心结果

1. 科研产出持续性规律：多数科学家科研生涯存在“产出高峰期”，且高峰期持续时长与早期科研启动速度、合作网络稳定性正相关。

2. 关键影响因素：机构支持力度、跨学科合作程度、早期科研成果认可度是影响科学家科研持续性的核心因素。

3. 群体差异：不同学科领域科学家的职业轨迹特征存在显著差异，基础学科科学家的科研产出持续性普遍高于应用学科。

四、结论与意义

结论：科学家科研轨迹受个体、机构、学科环境等多维度因素共同影响，构建长期追踪模型可有效预测科研产出潜力。

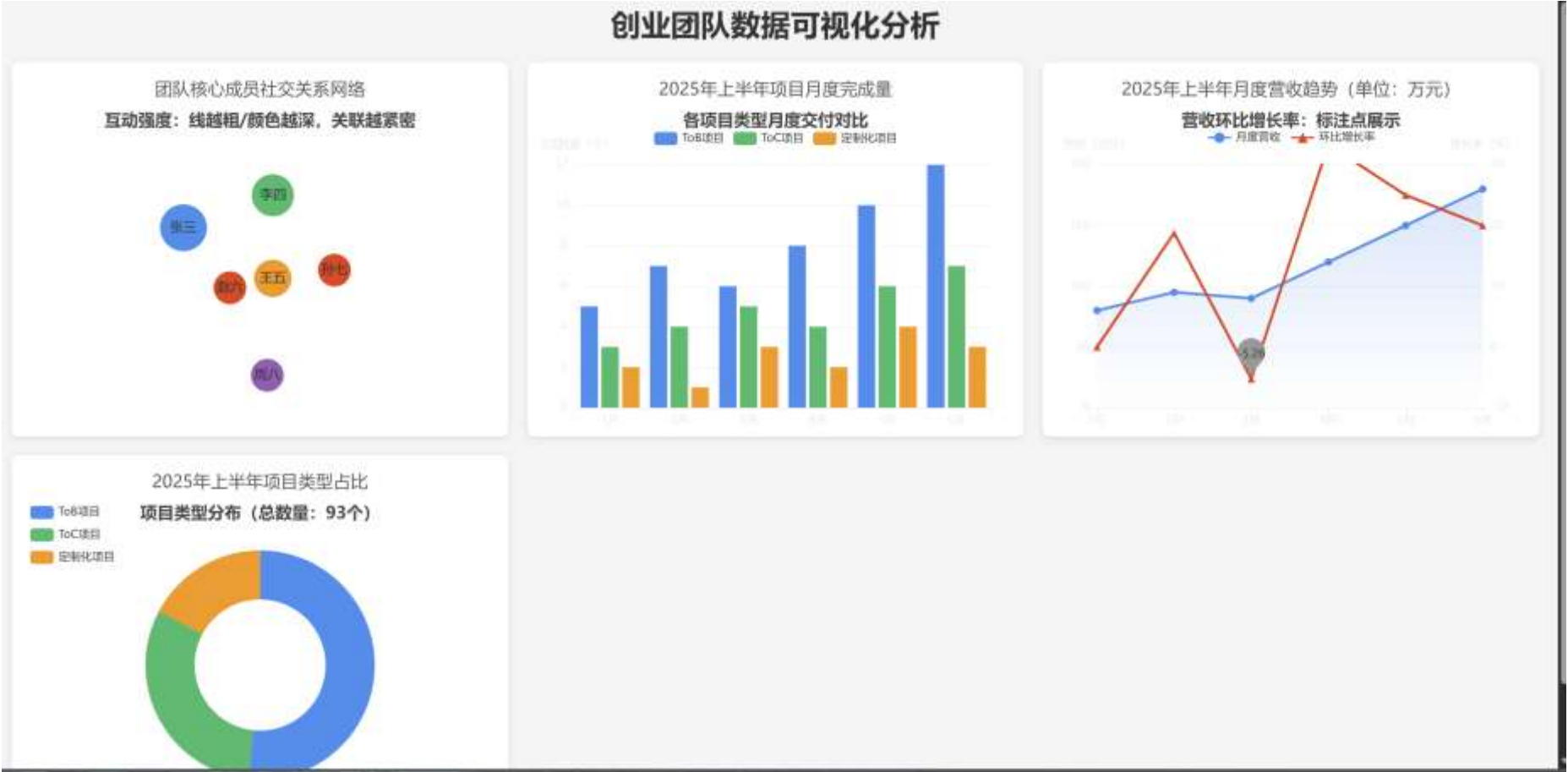
意义：为科研管理部门优化人才评价机制、制定科研支持政策提供数据支撑，助力科研生态的良性发展。

第三讲 词云与可视化

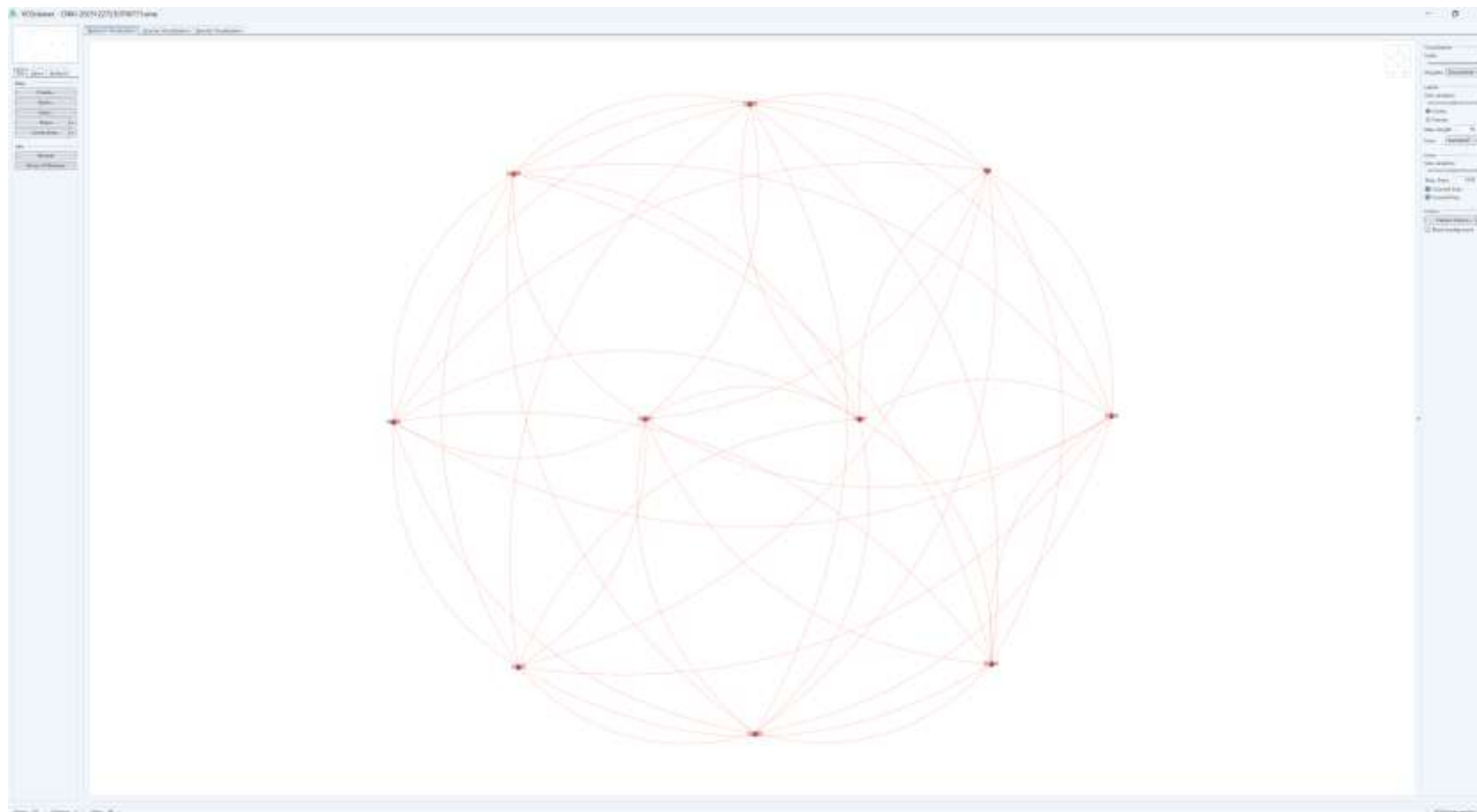
1.用任意一款词云工具，制作一个好看的词云（内容合理即可），并对词云图有一段话的解释。



2.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。



3.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。



4.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）。

```
1 # -*- coding: utf-8 -*-
2 """
3 词云图生成程序
4 """
5
6 # 导入需要的库
7 import matplotlib.pyplot as plt
8 from wordcloud import WordCloud, STOPWORDS
9
10 # 设置中文字体
11 plt.rcParams['font.sans-serif'] = ['SimHei']
12
13 # 读取文本文件
14 with open('scientist_words.txt', 'r', encoding='utf-8') as f:
15     text = f.read()
16
17 # 去除标点符号和数字
18 text = text.replace(' ', '').replace('.', '').replace(',', '').replace('!', '').replace('?', '').replace(';', '').replace(':', '').replace('(', '').replace(')', '').replace('%', '').replace('$', '').replace('&', '').replace('@', '').replace('#', '').replace('*', '').replace('~', '').replace('_', '').replace('-', '').replace(' ', '')
19
20 # 去除停用词
21 stopwords = set(STOPWORDS)
22
23 # 生成词云图
24 wc = WordCloud(font_path='simhei.ttf', width=1000, height=1000, background_color='white', max_font_size=200, min_font_size=10, random_state=42).generate(text)
25
26 # 显示词云图
27 plt.imshow(wc, interpolation='nearest')
28 plt.axis('off')
29 plt.title('科学家关键词词云图')
30 plt.show()
31
32 # 保存图片
33 plt.savefig('scientist_wordcloud.png')
```



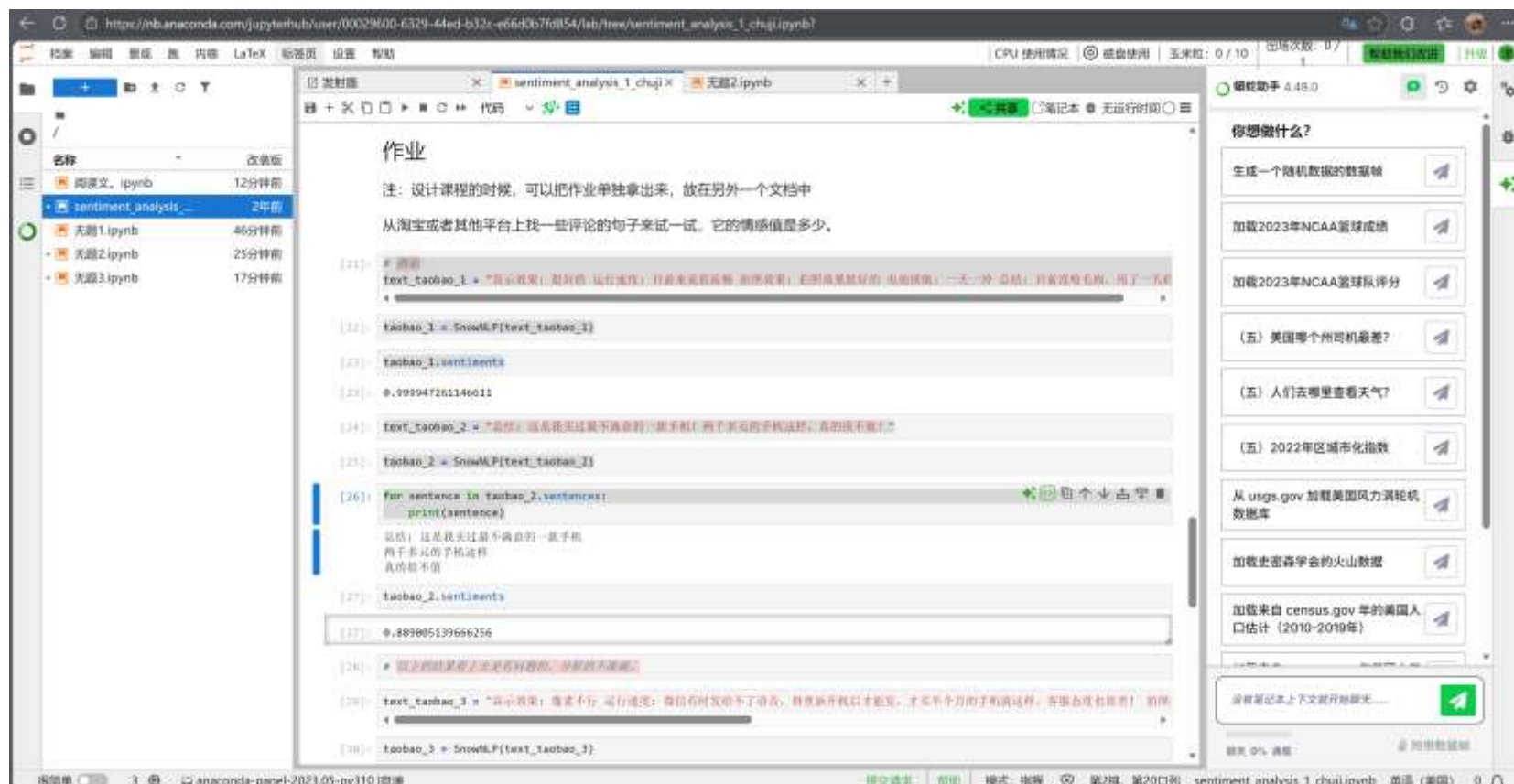
第四讲 情感分析

1.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；



2.完成sentiment_analysis_1-sentiment_analysis_4，4份代码。
做截图，并简要做代码运行总结分析。

1.



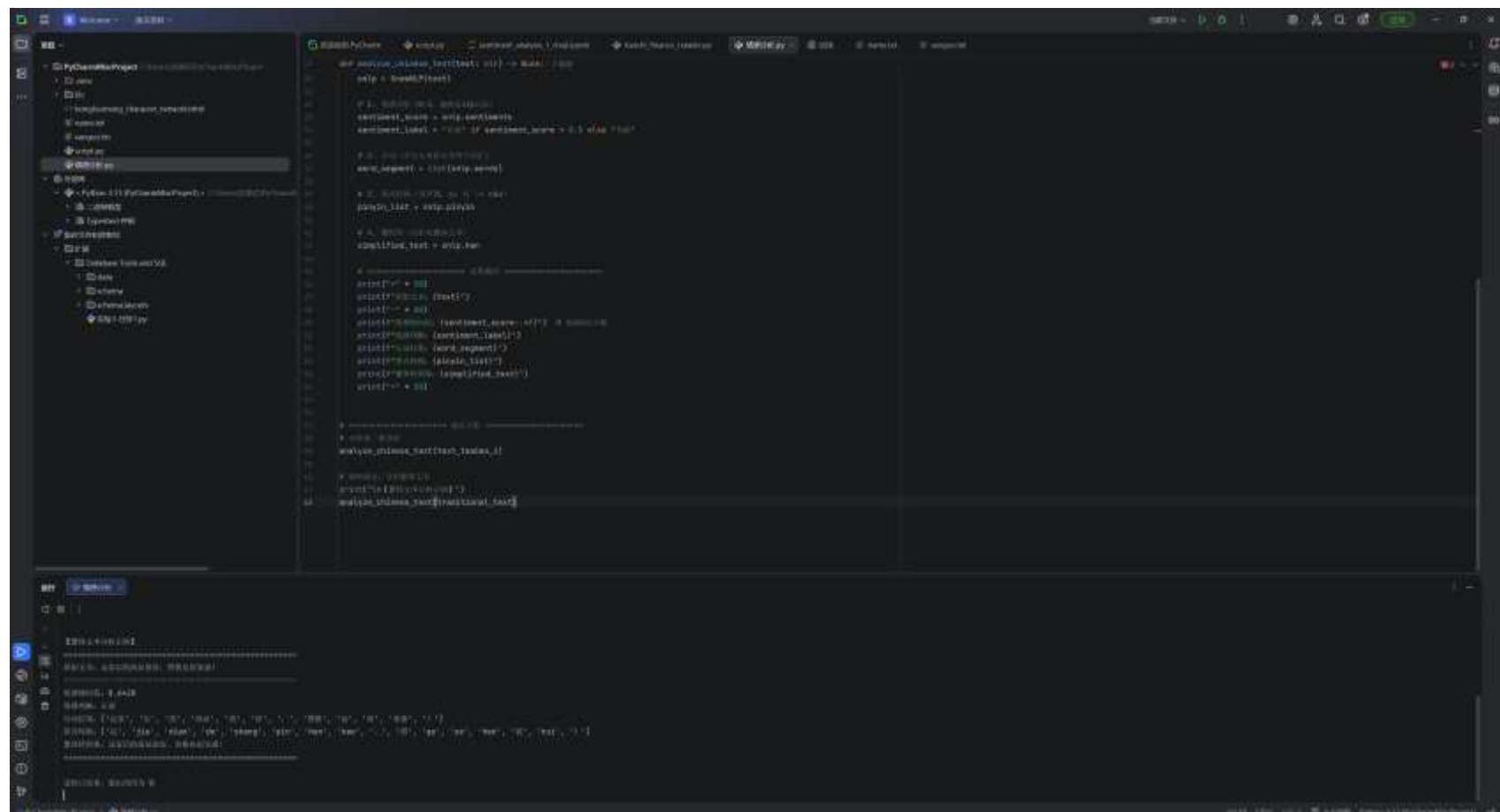
The screenshot displays a Jupyter Notebook titled "sentiment_analysis_1_chuji.ipynb" in a web browser. The notebook is divided into three main sections: a file browser on the left, a code editor in the center, and a sidebar on the right.

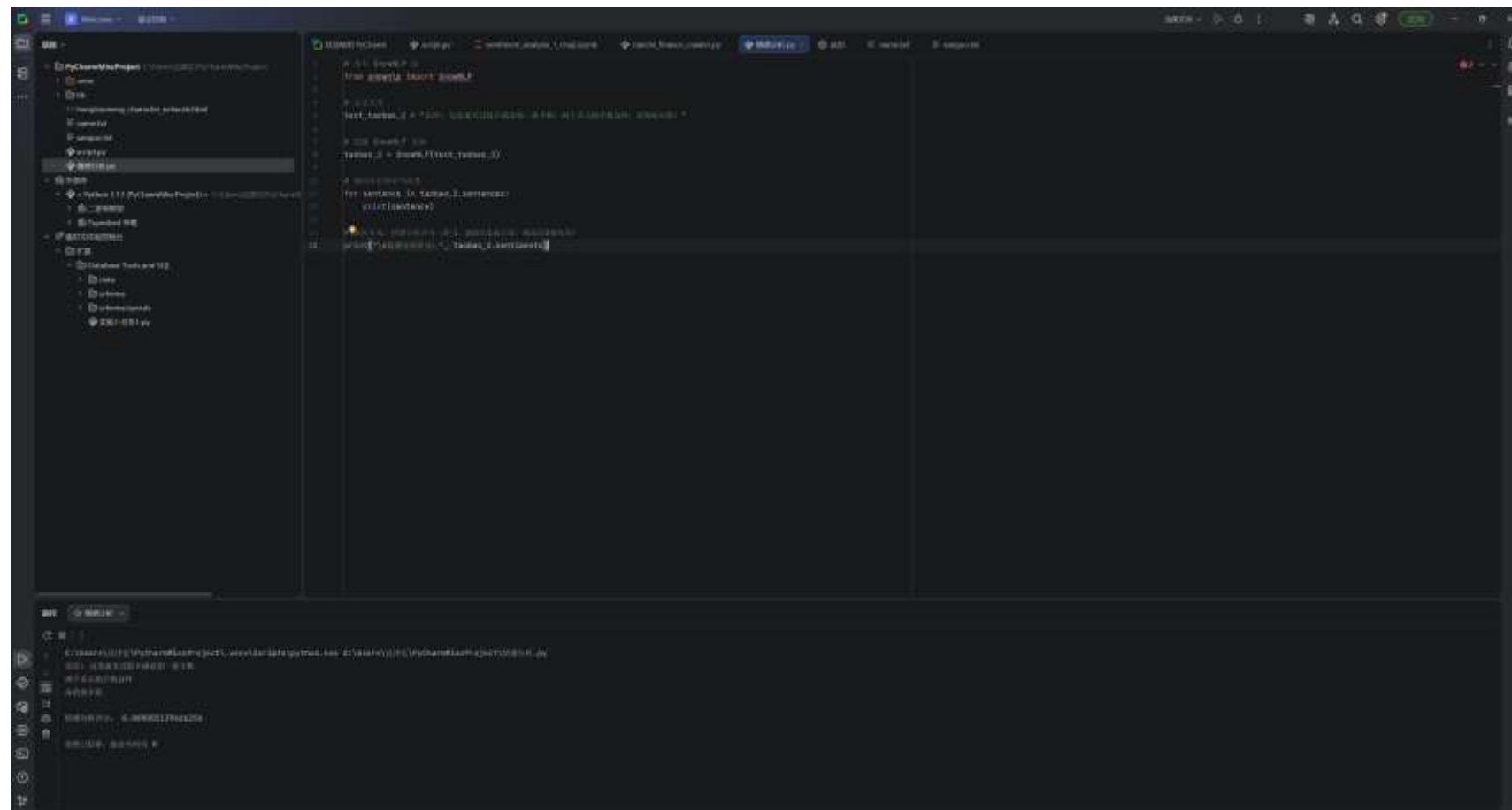
File Browser (Left): Shows a list of files including "同义词.ipynb", "sentiment_analysis_1.ipynb", "无题1.ipynb", "无题2.ipynb", and "无题3.ipynb".

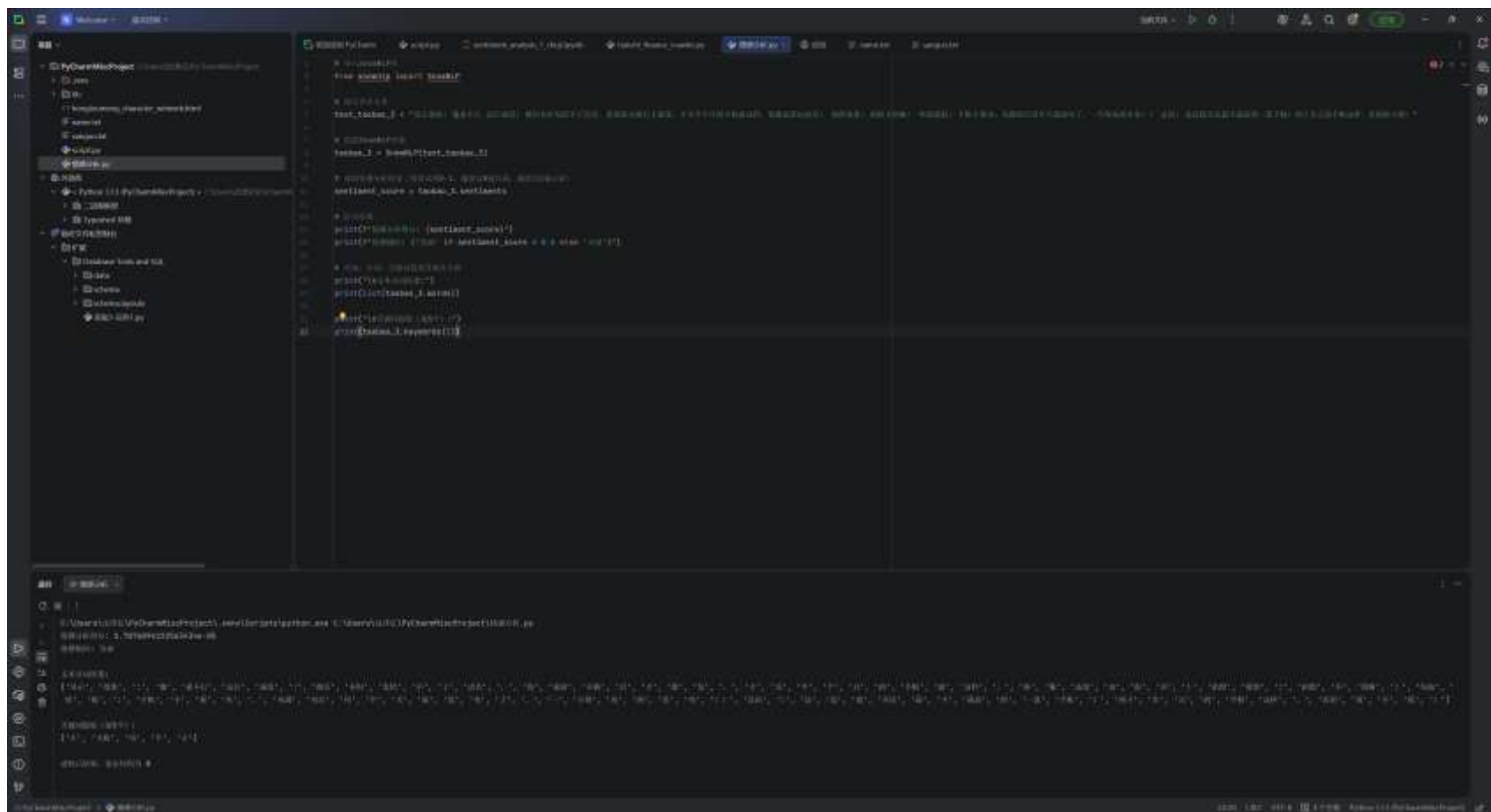
Code Editor (Center): Contains the following code cells:

- Cell 1:** A text input box with the text: "注：设计课程的时候，可以把作业单独拿出来，放在另外一个文档中。从淘宝或者其他平台上找一些评论的句子来试一试。它的情感值是多少。"
- Cell 2:** A code cell starting with `text_taobao_1 = "显示结果：(梨好色：运行速度：) 屏幕亮度高：(运行速度：) 电池续航：一天一冲：总结：对黄道吉日，用了一万..."`
- Cell 3:** A code cell with `taobao_1 = SnowMF(text_taobao_1)`
- Cell 4:** A code cell with `taobao_1.sentiments`
- Cell 5:** A code cell with `0.99947261146611`
- Cell 6:** A text input box with the text: "总结：这是我用过最不满意的一款手机！两千多买的手机这样真的不值！"
- Cell 7:** A code cell with `taobao_2 = SnowMF(text_taobao_2)`
- Cell 8:** A code cell with `for sentence in taobao_2.sentences: print(sentence)`
- Cell 9:** A code cell with `taobao_2.sentiments`
- Cell 10:** A code cell with `0.889085139696256`
- Cell 11:** A text input box with the text: "以上两结果上手机有问题，分析的不准确。"
- Cell 12:** A code cell with `text_taobao_3 = "显示结果：像是不行：运行速度：微信有时发不了语音，特意换手机以才发觉，才买半个月的手和就这样。客服态度也差！ 差评..."`
- Cell 13:** A code cell with `taobao_3 = SnowMF(text_taobao_3)`

Sidebar (Right): Contains a "你想做什么?" section with several buttons for data generation and analysis, such as "生成一个随机数据的数据库", "加载2023年NCAA篮球成绩", "加载2023年NCAA篮球评分", "(五) 美国哪个州司机最差?", "(五) 人们去哪里查看天气?", "(五) 2022年区域城市化指数", "从 usgs.gov 加载美国风力涡轮机数据库", "加载史密森学会的火山数据", and "加载来自 census.gov 年的美国人口估计 (2010-2019年)".







2.

The screenshot displays the Anaconda JupyterLab interface. On the left, a file browser shows a directory with files like 'README.ipynb', 'sentiment_analysis_...', and 'Untitled.ipynb'. The main area shows a notebook titled '情感分析-初级' (Sentiment Analysis - Beginner). The notebook content is in Chinese and includes a title, a link to a reference website, a paragraph about installing the environment, a code cell with a syntax error, and a section for features.

情感分析-初级

[参考网址在这里](#)

首先，需要安装情感分析的运行环境，也就是一些必要的程序包

- 在anaconda prompt里面运行

```
[2]: !pip install snowlp
!pip install -U textblob
!python -m textblob.download_corpora
```

Cell In[2], line 1
!pip install snowlp

SyntaxError: invalid syntax

英文

```
[3]: # 英文用TextBlob NLP处理包
```

Features

- Noun phrase extraction
- Part-of-speech tagging
- Sentiment analysis
- Classification (Naive Bayes, Decision Tree)
- Language translation and detection powered by Google Translate
- Tokenization (splitting text into words and sentences)
- Word and phrase frequencies
- Parsing
- n-grams
- Word inflection (pluralization and singularization) and lemmatization

Anaconda Assistant 4.48.0

What would you like to do?

The Anaconda Assistant is an AI-powered chat application designed to enhance the productivity of data scientists, developers, and researchers.

Get started

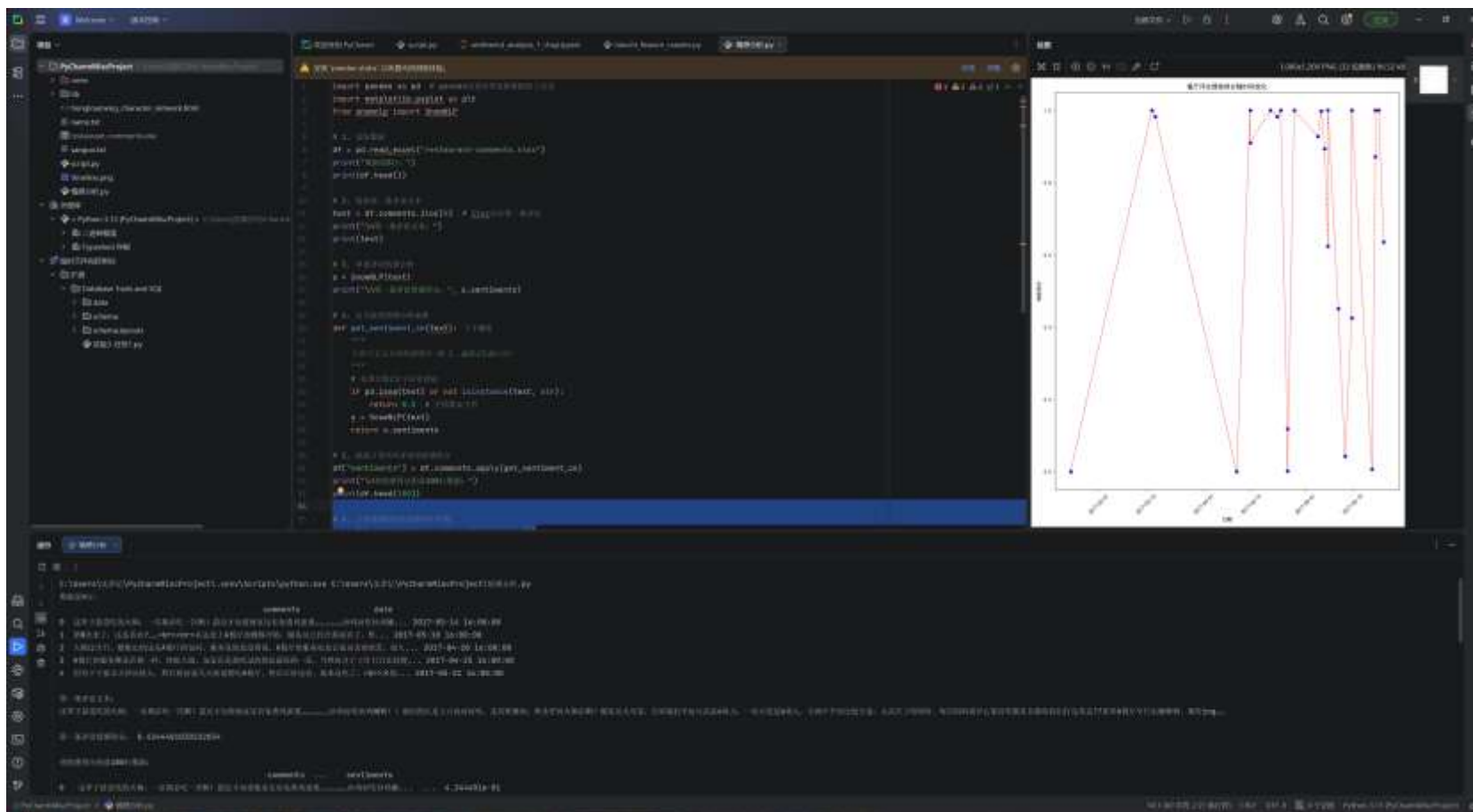
- Load volcano data from the Smithsonian Institute
- Plot the first 100 Fibonacci numbers
- Create a function to mask emails

Working with dataframes

- Load a DataFrame

Start chat without notebook context...

Chat: 0% full



3.

The screenshot displays the Anaconda Jupyter Notebook environment. The main workspace shows a notebook titled "sentiment_analysis_3_大模型_健康文本细粒度情感抽取.ipynb". The code cell contains the following text:

```
[ ]: # https://platform.deepseek.com/transactions
```

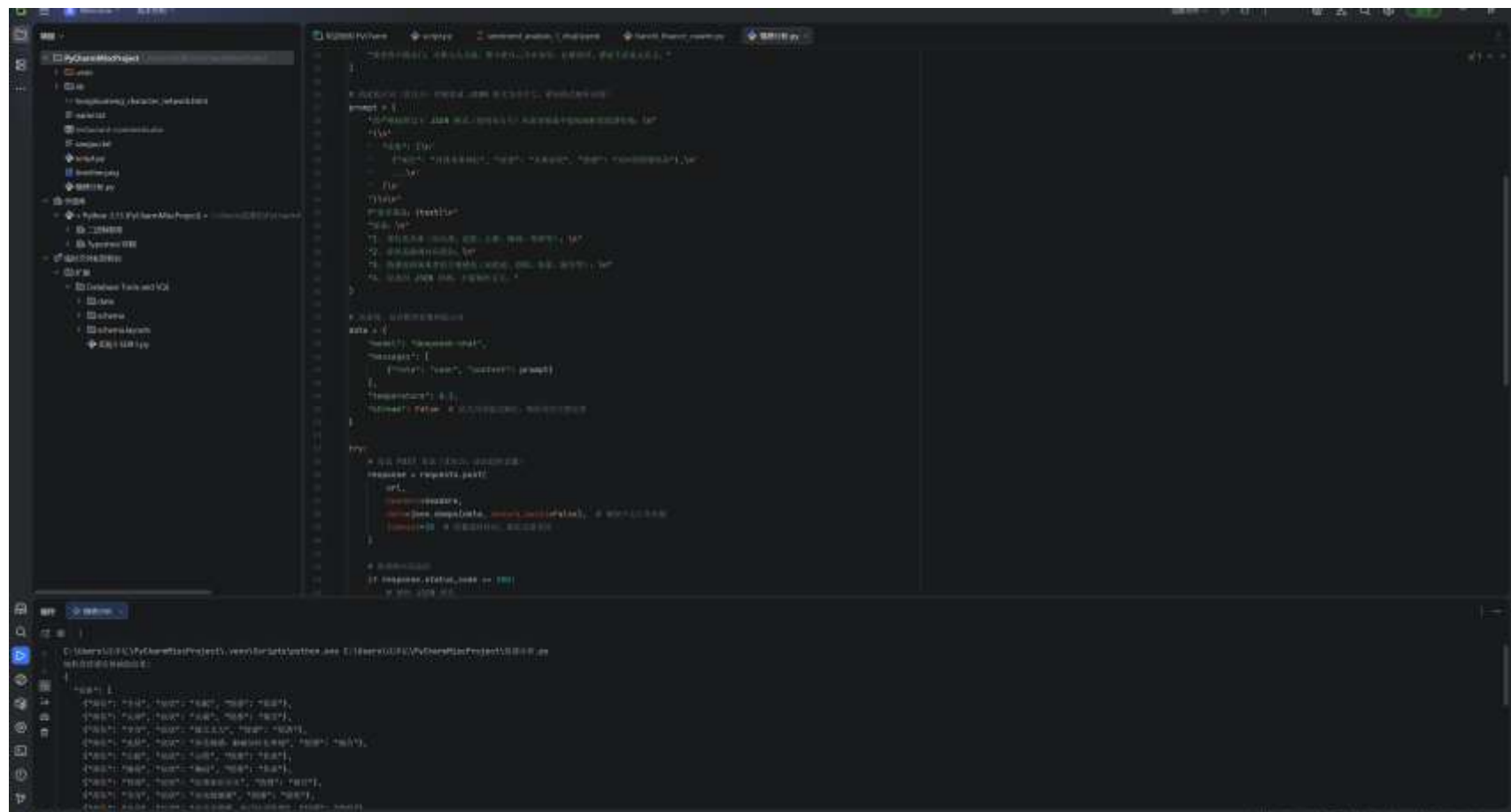
Below the code cell is a screenshot of the DeepSeek website, which shows a table of transactions. The table has columns for "Transaction ID", "Amount", "Status", and "Created At".

The left sidebar shows a file explorer with three notebooks: "sentiment_analysis_2_time", "sentiment_analysis_1_chuji", and "sentiment_analysis_3_大模型".

The right sidebar contains the Anaconda Assistant chat interface, which includes a "What would you like to do?" section with suggestions like "Load volcano data from the Smithsonian Institute", "Plot the first 100 Fibonacci numbers", and "Create a function to mask emails". It also has a "Get started" section with a "Load a DataFrame" button and a "Working with dataframes" section with a "Load a DataFrame" button.

At the bottom of the notebook, there is a code cell with the following text:

```
[1]: import requests
import json
```



```
C:\Users\吴建亿>PyCharmMacProject\venv\Scripts\python.exe C:\Users\吴建亿
PyCharmMacProject\情感分析.py /
```

按照度情感实体抽取结果：-

[-

 "实体":[-

 {"部位":"全身","症状":"失眠","情感":"低落"},

 {"部位":"头部","症状":"头痛","情感":"痛苦"},

 {"部位":"全身","症状":"疲乏无力","情感":"憔悴"},

 {"部位":"皮肤","症状":"异常敏感,触碰如针扎疼痛","情感":"痛苦"},

 {"部位":"心脏","症状":"心慌","情感":"焦虑"},

 {"部位":"胸部","症状":"胸闷","情感":"焦虑"},

 {"部位":"背部","症状":"沉重如压石头","情感":"痛苦"},

 {"部位":"全身","症状":"对光线敏感","情感":"恐惧"},

 {"部位":"全身","症状":"对声音敏感,电话铃声烦躁","情感":"恐惧"},

 {"部位":"全身","症状":"不愿出门,不与人交流,关在屋里拉窗帘","情感":"低落"},

]-

 {"部位":"全身","症状":"感觉生活毫无意义","情感":"低落"}]

]-

]-

]-

按照层的结构化数据：-

部位：全身 | 症状：失眠 | 情感：低落-

部位：头部 | 症状：头痛 | 情感：痛苦-

部位：全身 | 症状：疲乏无力 | 情感：憔悴-

部位：皮肤 | 症状：异常敏感,触碰如针扎疼痛 | 情感：痛苦-

部位：心脏 | 症状：心慌 | 情感：焦虑-

部位：胸部 | 症状：胸闷 | 情感：焦虑-

1. 请问您觉得大语言模型在识别患者的身体、心理、情感的工作中，表现如何？

2. 除了医疗健康领域的情感理解与识别，您还想到哪些其他重要的领域，可以开展类似的工作呢？

除健康领域外，以下领域对细粒度情感识别需求迫切，且落地价值显著。具体如下：

一、智能客服与用户运营

核心应用场景

1. 实时识别客户咨询中的愤怒、失望等情绪，分级响应。

2. 分析售后评论的情感实体（如“屏幕卡顿→故障”），定位产品痛点。

价值与目标

提升客户满意度，缩短问题解决周期，降低产品迭代。

二、教育与学习分析

核心应用场景

1. 分析学生作业/课堂发言中的情感（如“公式难懂→挫败”），定制学习路径。

2. 评估在线课程的情感反馈，优化教学内容。

价值与目标

实现个性化教学，降低辍学率，提升学习体验。

三、舆情监测与品牌管理

核心应用场景

1. 监测社交媒体中品牌相关的情感实体（如“新机跑一公里”）。

2. 预警负面舆情扩散，快速响应危机事件。

价值与目标

维护品牌形象，及时调整公关策略，量化市场情绪。

四、司法与纠纷

核心应用场景

1. 分析证人证言中的情感波动（如“目击案发时间→紧张”），辅助可信度判断。

2. 识别诉讼文本中的焦虑、恐惧等情绪，优化诉讼策略。

价值与目标

提升案件处理效率，保障司法公正（需结合人工审核）。

五、智能导航与车载交互

核心应用场景

1. 识别驾驶员的烦躁、疲惫等情绪，触发安全提醒（如“疲劳驾驶→警告”）。

2. 调整车内环境（如音乐、温度）适配情绪。

价值与目标

降低交通事故风险，提升驾乘体验。

六、内容创作与媒体

核心应用场景

1. 分析影视剧广告的情感节奏（如“剧情转折→惊喜”），优化内容结构。

2. 生成符合目标情感的文章（如公益广告的情感文案）。

价值与目标

提升内容传播力，增强用户共鸣。

七、心理健康与社会服务

核心应用场景

1. 社区/学校的心理筛查（如学生日记中的“孤独→预警”）。

2. 老年陪伴机器人的情感响应。

价值与目标

扩大心理健康资源覆盖，早期干预风险。

八、金融风控

核心应用场景

1. 分析贷款申请文本中的焦虑、隐瞒等情绪，辅助信用评估。

2. 监测交易对话中的异常情绪，预警欺诈风险。

价值与目标

降低坏账率，提升风控精度。

”

4.

The screenshot displays the Anaconda Assistant 4.48.0 interface. The main window is a Jupyter Notebook titled 'sentiment_analysis_2_time'. The notebook content includes a title '文学作品的情感轨迹' (Emotional Trajectory of Literary Works) and a list of steps for sentiment analysis. Below the text is a code cell with Python code for reading a file, splitting it into sentences, and printing them. The sidebar on the right contains a section 'What would you like to do?' with several suggested actions like 'Load volcano data from the Smithsonian Institute' and 'Plot the first 100 Fibonacci numbers'. At the bottom, there is a chat input area with the text '41: 到这时, 我赶紧去撸他。'.

File Edit View Run Kernel LaTeX Tabs Settings Help CPU Usage Disk Usage Kernels: 1 / 10 Apps: 0 / 1 Help Us Improve Upgrade

sentiment_analysis_2_time X sentiment_analysis_4_大概! X sentiment_analysis_1_chuji X sentiment_analysis_3_大概! X

Code Share Notebook No Runtime

Name Modified

- sentiment_analysis_... yesterday
- sentiment_analysis_... yesterday
- sentiment_analysis_... now
- sentiment_analysis_... now

文学作品的情感轨迹

- 步骤是这样的:
- 1) 读取txt文本, 按照句号、感叹号, 对文本进行分句, 并且对句子进行编号;
- 2) 使用大语言模型, 提取以上编号句子, 每句话中的情感词, 或者隐含的情感实体;
- 3) 用大语言模型计算出, 每一个情感词或者情感词所在句子的情感得分;
- 4) 然后绘制一个时间序列的可视化图, 横坐标是句子的序号, 纵坐标是情感得分, 每一个散点是句子对应的情感词, 情感词作为散点的标签, 然后用虚线把这些情感词连起来。

```
[1]: import re

with open("资源.txt", "r", encoding="utf-8") as f:
    text = f.read()

# 按照句号、感叹号进行分句 (忽略换行)
sentences = re.split(r'(?<=[。!])', text)
sentences = [s.strip() for s in sentences if s.strip()]
numbered_sentences = [(i + 1, s) for i, s in enumerate(sentences)]
# numbered_sentences = numbered_sentences[0:5] # 大概截取前5句

for num, sent in numbered_sentences:
    print(f"{num}: {sent}")
```

41: 到这时, 我赶紧去撸他。

What would you like to do?

The Anaconda Assistant is an AI-powered chat application designed to enhance the productivity of data scientists, developers, and researchers.

Get started

- Load volcano data from the Smithsonian Institute
- Plot the first 100 Fibonacci numbers
- Create a function to mask emails

Working with dataframes

- Load a DataFrame

Start chat without notebook context...

Chat 0% full Attach DataFrame

代码运行总结分析

1. 核心功能模块

文本情感分析: 通过 SnowNLP 库实现中文文本的情感极性 (正面 / 负面)、情感得分计算。同时支持文本分词、关键词提取。

批量数据处理: 结合 pandas 读取 CSV 格式的评论数据, 对批量文本进行情感分析并生成结果列。

情感可视化: 通过 matplotlib 绘制情感得分的折线图, 直观展示文本情感的波动趋势。

大模型接口调用: 对接 deepseek-chat 模型的 API, 实现文本的结构化解析 (如提取人物、情绪、行为等要素)。

2. 运行状态分析

基础功能 (SnowNLP 部分): 从输出结果看, 文本分词、情感得分计算、关键词提取均正常运行, 能正确输出分词列表、情感得分 (如 "0.6238") 和关键词集合。但 SnowNLP 本身是轻量级工具, 情感分析的精度可能有限 (比如部分文本的情感极性判断可能存在偏差)。

批量处理与可视化: 成功读取 CSV 数据并批量计算情感得分, 且可视化图表正常生成 (折线图展示了情感得分的变化), 这部分流程是流畅的。

大模型 API 调用: 代码中尝试通过 requests 调用 deepseek-

chat 接口, 输出结果显示已成功提取文本中的人物、情绪、行为等要素 (如 "情绪: 0.7""行为: 道歉"), 接口通信和解析逻辑是有效的。

3. 潜在优化点

情感分析精度: SnowNLP 的效果较弱, 可替换为更精准的工具 (如 TextBlob 的中文扩展、百度 / 腾讯的 NLP 接口)。

API 稳定性: 大模型调用时可增加异常处理 (如网络超时、接口报错的重试逻辑)。

可视化增强: 可优化图表的标注 (如添加横轴标签、图例), 让情感趋势更易读。

="

第六讲 知识图谱理念

1. 实际产业案例分析：使用3-5页PPT对“阿里商品大脑”、“美团大脑”、“丁香医生知识图谱”、“领英知识图谱”…其中任意一家机构/公司最新的知识图谱生态构建，进行简要介绍与分析。要求：需要是最新进展（不能复制课程PPT中的内容）；可以是一个简单的案例；有自己的评价。自由发挥。（营销、信管，都可以结合自己的专业兴趣，自由选择分析对象）



**AI重构20亿商品——阿里商品大脑
知识图谱生态升级实践：电商知识
图谱的产业落地与价值赋能**



目录

01

阿里商品大脑知识图谱生态概况
(2025最新进展)

02

核心技术支撑与生态构建逻辑

03

案例落地：天猫双11的营销与运营赋能

04

评价与展望

生态核心定位



阿里商品大脑知识图谱生态以“全链路智能提效”为核心，致力于构建覆盖“人-货-场”的电商知识图谱基础设施。在过去，电商运营主要依赖流量分发模式，而如今，阿里通过知识图谱的深度构建，实现了从“流量分发”到“精准价值匹配”的底层升级。这一转变意味着，它不再仅仅是将商品信息广泛撒网式地推送给用户，而是通过对用户需求和商品特性的精准分析，实现两者的高效匹配。例如，在天猫双11等核心场景中，它作为智能引擎，能够根据消费者过往的购物偏好、浏览记录以及实时搜索行为，精准推送符合其需求的商品，大大提升了购物效率和用户满意度。





2025关键升级



商品理解重构

基于大模型技术，阿里完成了平台20亿商品的深度语义解析与标签体系重构。在传统模式下，商品信息标注主要依靠人工，不仅效率低下，而且容易出现疏漏和不准确的情况。而现在，借助大模型强大的学习和分析能力，商品信息丰富度平均提升34%。以宠物行业为例，对于猫粮，它能够精准解析出成分、功效等维度的标签，如明确标注出含有哪些营养成分，对猫咪的毛发、肠胃等有何种功效，让消费者在选购时能更全面地了解商品信息，做出更合适的选择。



生态边界拓展

阿里商品大脑知识图谱生态从单一商品知识管理，成功延伸至商家经营全链路。以往，各个业务环节的数据相互独立，形成了数据孤岛，无法实现协同效应。如今，通过整合搜索、推荐、营销、客服、物流等多环节数据，形成了闭环式知识服务生态。商家可以根据搜索数据了解消费者的热门需求，结合推荐数据优化商品展示策略，利用营销数据制定更具针对性的促销活动，借助客服数据改进商品和服务，依据物流数据优化配送流程，从而实现全方位的经营优化。



技术架构升级

融合多模态知识融合技术是阿里的又一重要升级。在复杂的电商场景中，商品信息以多种形式呈现，包括文本描述和图像展示等。通过这种融合技术，将图像属性识别与文本描述匹配准确率提升至89.3%。比如，当消费者搜索一款蓝色的纯棉衬衫时，系统能够快速准确地将衬衫的蓝色图像属性与“蓝色”“纯棉”等文本描述进行匹配，即使商品描述存在语义模糊或不完整的情况，也能实现精准匹配，为消费者提供更准确的搜索结果。

核心技术底座

多模态知识抽取

阿里整合文本（商品详情、评价）、图像（商品外观、属性）等异构数据，通过千亿参数预训练模型实现实体识别F1值87%+。商品详情中的文字描述和用户评价中的反馈，以及商品图片展示的外观和属性细节，都是重要的信息来源。通过这种多模态知识抽取技术，能够更全面、准确地提取商品的关键信息，为后续的知识图谱构建和应用提供坚实的数据基础。例如，对于一款手机，不仅能从文本中提取出品牌、型号、配置等信息，还能从图像中识别出手机的颜色、外观设计特点等，从而构建出更完整的商品知识。

动态知识推理引擎

动态知识推理引擎支持上下位推理、不一致性校验等多场景推理。在商品发布过程中，它能发挥重要作用，如自动拦截“产地标注矛盾”商品。当一款商品同时标注来自两个不同产地时，推理引擎能够迅速识别出这种矛盾信息，并在发布端实现毫秒级实时审核，避免错误信息进入市场，保障了商品信息的准确性和可靠性，维护了电商平台的良好秩序。

全链路数据融合

阿里打通淘宝、天猫、菜鸟、支付宝等多业务数据，构建包含百亿级三元组的知识网络，覆盖商品、品牌、用户、物流等9大类核心本体。这些业务数据分别从不同角度反映了电商生态的各个环节，通过全链路数据融合，能够形成一个全面、立体的知识网络。比如，通过融合淘宝和天猫的商品销售数据、菜鸟的物流配送数据以及支付宝的支付数据，可以深入了解用户的购物行为、商品的流通过程以及支付习惯等，为电商运营和决策提供丰富的数据支持。

生态构建逻辑

阿里以“商品知识标准化”为基础，向上赋能消费者端需求匹配，向下支撑商家端经营决策。在消费者端，通过知识图谱实现“需求识别- 流量匹配- 交易转化- 履约服务”的全链路智能协同。当消费者在平台上搜索商品时，系统能够准确识别其需求，通过知识图谱将需求与合适的商品进行流量匹配，促进交易转化，并在履约服务环节确保商品能够及时、准确地送达消费者手中。在商家端，知识图谱为商家提供了全面的市场信息和消费者洞察，帮助商家制定合理的经营策略，优化商品选品、定价和营销活动，从而提升经营效率和竞争力。

核心应用场景

消费者端：精准需求匹配

基于知识图谱的“AI万能搜”实现复杂语义理解，搜索相关性提升20个百分点。在以往的搜索中，消费者可能因为输入的关键词不够准确或完整，导致搜索结果不理想。而现在，“AI万能搜”能够理解消费者的自然语言表达，即使是模糊的需求，也能准确匹配到相关商品。比如，消费者输入“适合跑步时穿的舒适鞋子”，它能快速筛选出符合要求的跑鞋，并根据消费者的偏好和历史记录进行排序推荐。同时，推荐信息流点击量提升10%，从“猜你喜欢”升级为“懂你所需”，真正解决了模糊需求检索的痛点，为消费者提供了更个性化、精准的购物体验。

商家端：智能经营赋能

通过“AI店长+6个AI员工”的Agent组合模式，为500万商家提供定制化经营策略。以“货品全站推”为例，依托知识图谱实现广告ROI提升12%，预售期间商家ROI同比增幅超15%。在预售活动中，商家可以根据知识图谱提供的消费者需求和市场趋势信息，精准定位目标客户，制定更有针对性的预售策略，从而提高投资回报率。店小蜜5.0客服基于商品知识图谱实现90%常规咨询自动解决，成交转化贡献占比达25%。这大大减轻了商家的客服压力，提高了服务效率，同时也促进了成交转化，为商家带来了更多的商业机会。






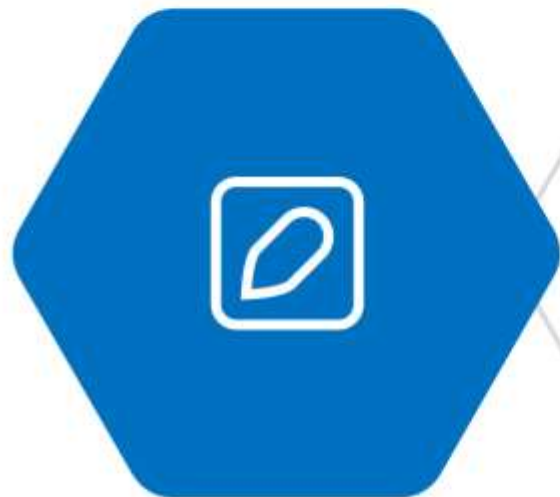
量化价值成果



阿里商品大脑知识图谱生态在天猫双11中的应用，带来了显著的量化价值成果。流量匹配效率实现双位数增长，这使得商品能够更精准地触达目标消费者，提高了流量的利用效率。这种高效的流量匹配带动了天猫双11开卖首小时80个品牌成交破亿，30516个品牌成交翻倍，充分验证了知识图谱生态在电商领域的强大商业价值，展示了其对电商业务增长的巨大推动作用。



核心优势评价



场景落地能力强

阿里商品大脑知识图谱生态精准切入电商核心痛点，将知识图谱从单纯的技术工具转变为电商运营的基础设施。它紧密围绕电商业务流程，从商品管理、搜索推荐到交易转化等各个环节，都能发挥实际作用，直接关联商业转化成果。例如，通过优化搜索和推荐算法，提高了商品与消费者需求的匹配度，从而促进了销售增长，切实解决了电商行业长期存在的效率低下和用户体验不佳等问题。



生态协同价值高

该生态打通了“人-货-场”全链路数据，打破了数据孤岛，实现了知识的高效复用与赋能。不同业务环节的数据相互融合，使得各个环节能够协同工作，形成一个有机的整体。比如，消费者数据可以帮助商家优化商品选品和营销活动，商家数据又能为消费者提供更个性化的服务，这种协同效应提高了整个电商生态系统的运行效率，符合信息管理的协同化趋势。



商业价值闭环

阿里商品大脑知识图谱生态既提升了消费者体验，又降低了商家经营成本，形成了“技术-效率-增长”的正向飞轮。在消费者端，通过精准的商品推荐和便捷的购物服务，提高了购物满意度和忠诚度；在商家端，借助智能经营策略和自动化工具，节省了约30%的人工工作量，降低了运营成本。这种商业价值闭环促进了电商业务的可持续发展，为阿里在电商市场的竞争中赢得了优势。

现存挑战

跨领域知识迁移难题

不同行业商品知识体系差异大，这给阿里商品大脑知识图谱生态带来了跨领域知识迁移的难题。通用模型在适配不同行业时，适配率仍有提升空间，当前行业标准化产品适配率不足40%。比如，时尚行业注重款式、风格等因素，而电子产品行业更关注性能、参数等指标，如何让通用模型能够准确理解和处理不同行业的复杂知识，是需要解决的问题，这关系到知识图谱生态在更广泛领域的应用和拓展。

中小商家接入门槛

复杂知识服务工具对于中小商家来说，存在接入门槛较高的问题。中小商家通常资源有限，技术能力相对较弱，难以充分利用这些复杂的工具。因此，复杂知识服务工具的轻量化适配需进一步优化，使其更易于中小商家上手和使用，从而提升中小商家在电商生态中的参与度和竞争力，促进整个电商生态的均衡发展。

未来展望

深化跨境场景布局

依托阿里国际站资源，阿里计划构建多语言跨境商品知识图谱。随着全球贸易的不断发展，跨境电商市场潜力巨大。通过构建多语言跨境商品知识图谱，能够更好地满足不同国家和地区消费者的需求，实现全球采购场景的精准撮合。例如，为不同语言背景的消费者提供准确的商品信息和推荐服务，帮助商家拓展国际市场，提升阿里在全球电商市场的影响力。

融合隐私计算技术

在合规前提下，阿里将融合隐私计算技术，实现跨机构数据知识融合。数据是知识图谱生态的重要支撑，但在数据共享和融合过程中，隐私保护至关重要。通过融合隐私计算技术，能够在确保数据安全和隐私的基础上，进一步提升知识的广度与深度。比如，与金融机构、物流企业等跨机构进行数据合作，整合多方数据资源，为电商运营提供更全面、深入的数据分析和决策支持。

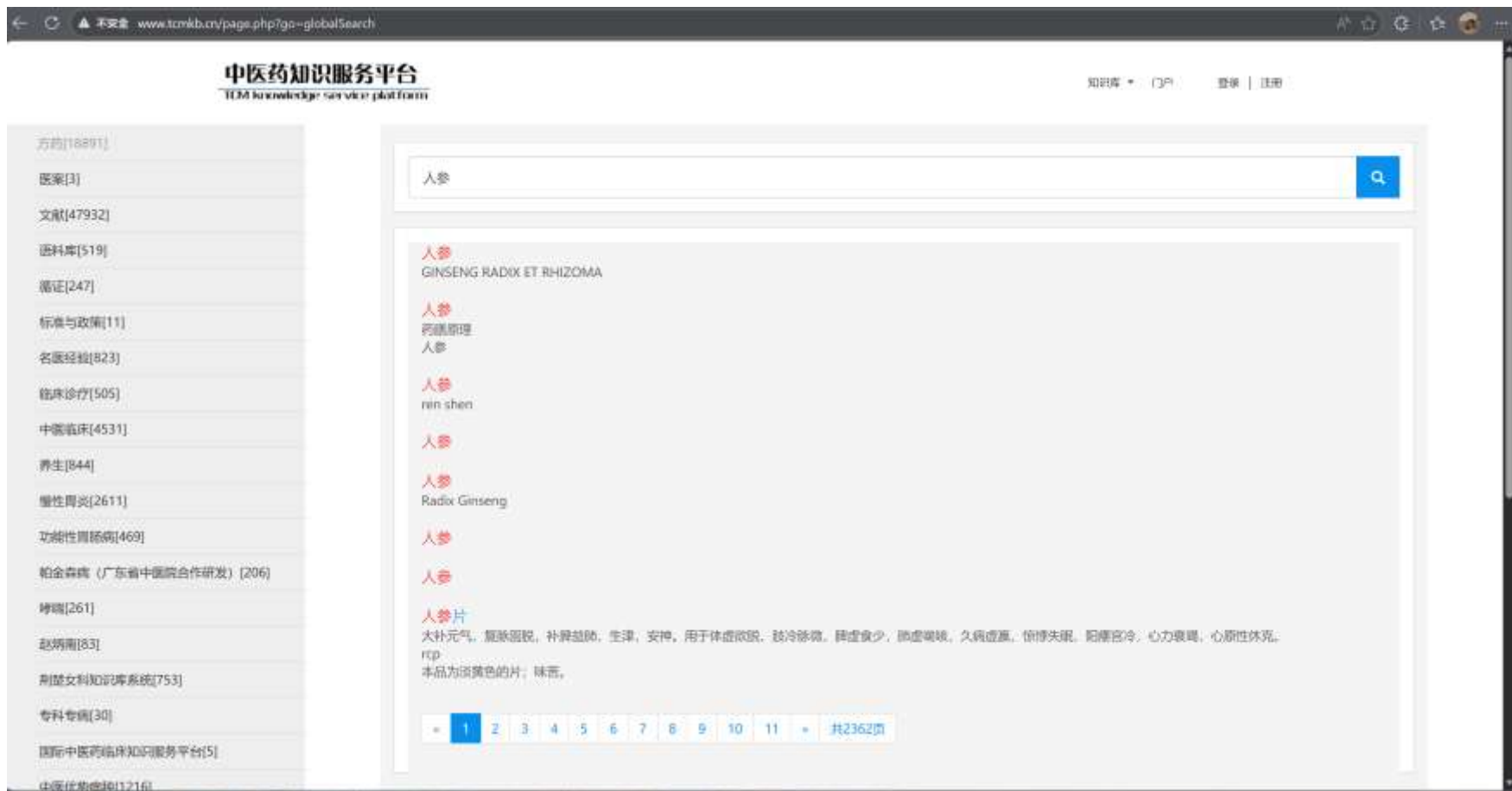
轻量化赋能中小商家

推出标准化知识服务SaaS工具是阿里未来的重要举措之一。中小商家在电商生态中占据着重要地位，但由于自身资源和技术限制，往往难以享受到优质的知识服务。通过推出标准化知识服务SaaS工具，能够降低中小商家的使用门槛，使其能够便捷地获取所需的知识服务，提升经营效率和竞争力，促进电商生态的繁荣发展。

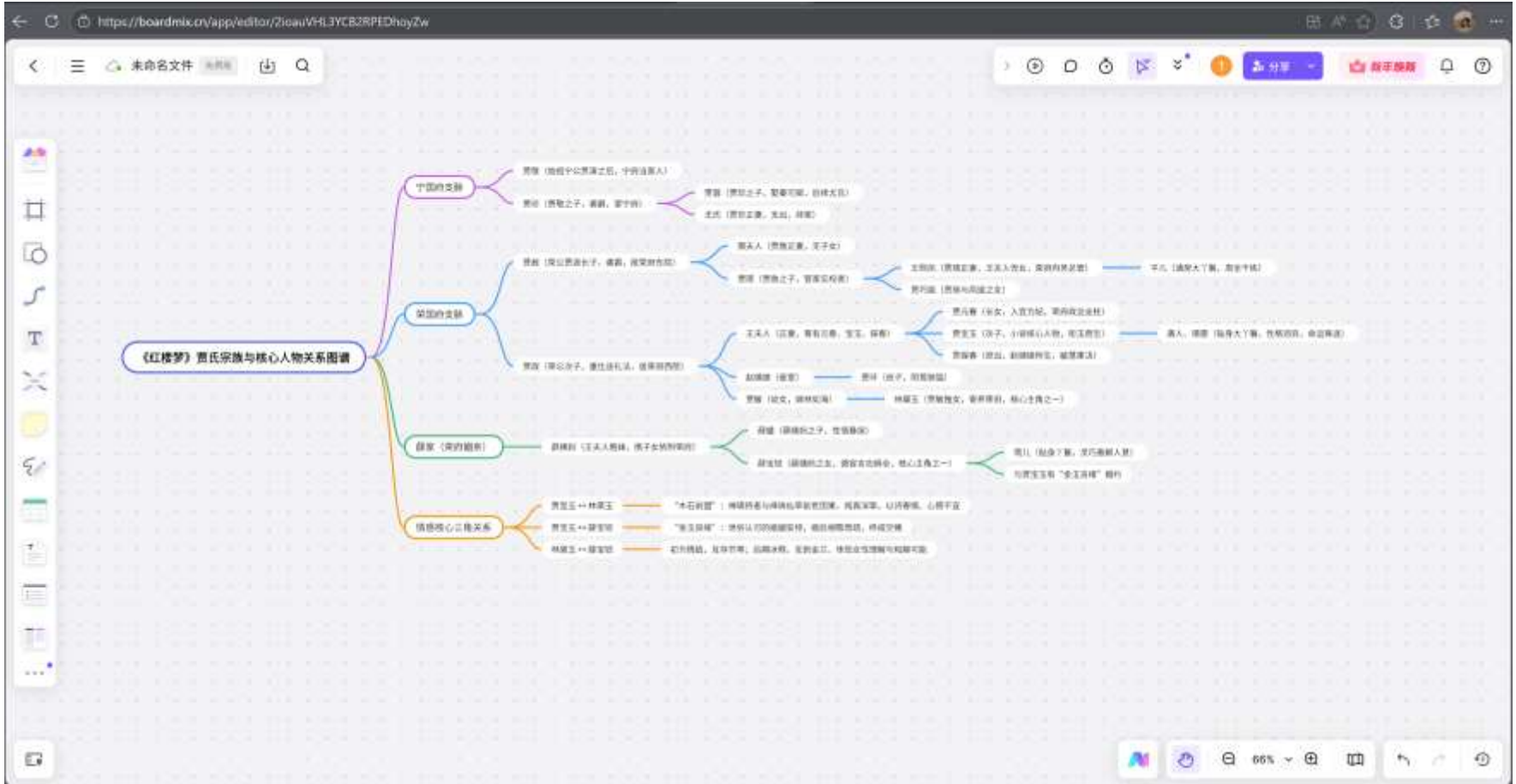


第六讲 (2) 知识图谱工具

1.使用PPT中知识图谱链接平台，检索、截图（大词林等，可用的）；



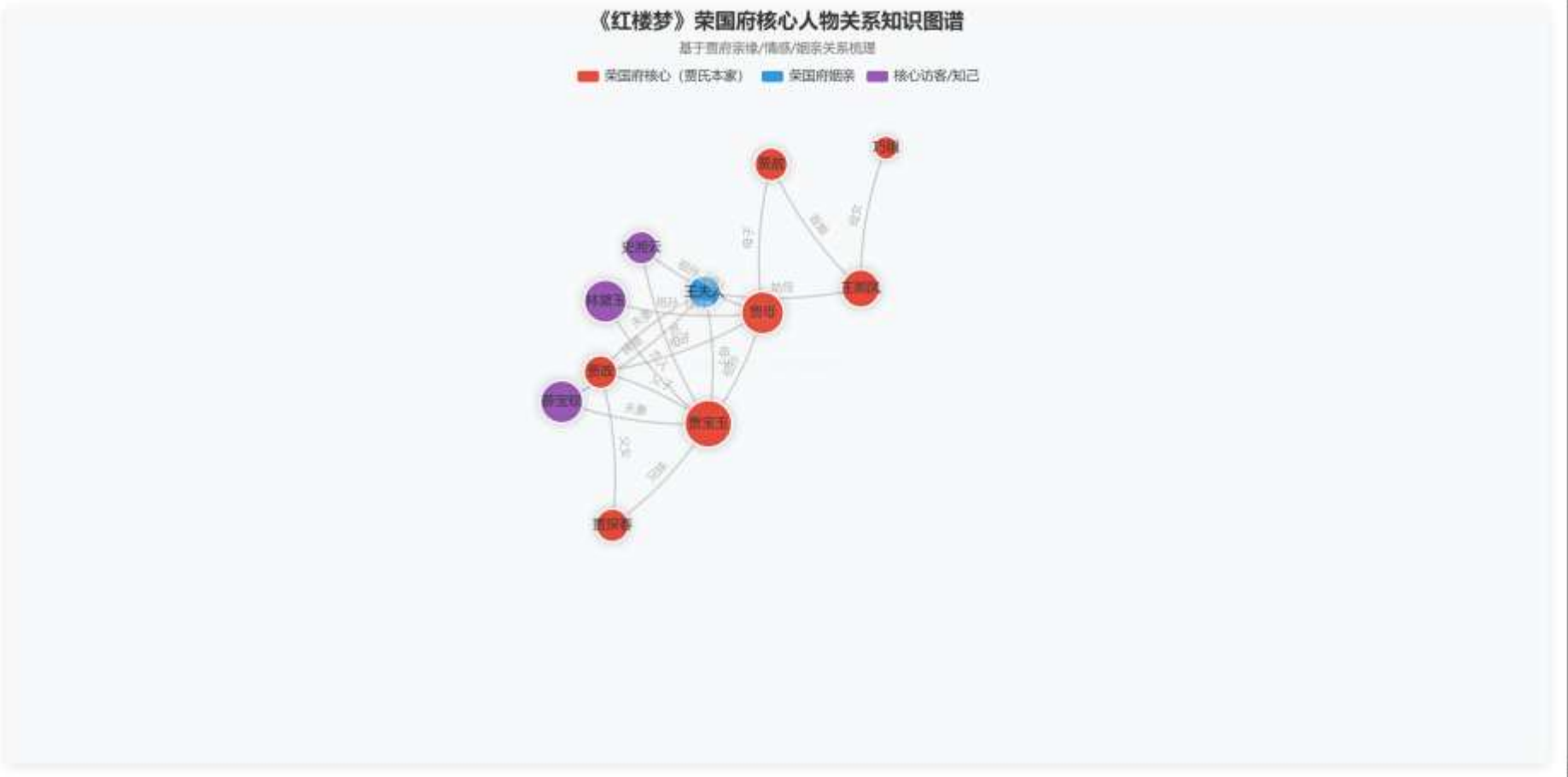
2.使用白板建模绘制一个你感兴趣的“知识图谱”，可以是人物关系，也可以是事物关系，或者概念之间的关系等等，并解释你绘制的图谱；



图谱关键关系解释

1. 宗族核心：以宁荣二公为始祖，分为宁国府（贾敬支脉）和荣国府（贾赦、贾政支脉）两大支，荣国府是全书故事的主要发生地。
2. 姻亲网络：通过 ** 王夫人（王家）、薛姨妈（王家）、王熙凤（王家）、林黛玉（林家）** 的婚姻关系，将贾、王、薛、林四大家族紧密连接，体现了“一荣俱荣，一损俱损”的家族命运。
3. 情感主线：贾宝玉、林黛玉、薛宝钗的“木石前盟”与“金玉良缘”是全书的核心情感冲突，红色虚线清晰标注了三人的情感关系。
4. 丫鬟群体：以袭人、晴雯、紫鹃、莺儿为代表的丫鬟，是核心人物的贴身助手，也是故事中重要的配角，用黄色椭圆区分，体现了等级差异。

3.使用echarts中的关系图，绘制作业2）中的“知识图谱”



4.使用Neo4j（可在线版本），编程绘制一款（简单）知识图谱（内容不限）（仅信管）

