

第四届“卿云杯”全国通识课程论文大赛

封面样张

学校	上海财经大学	院系	会计学院
专业	会计学	姓名	李一凡
年级	大一	任课教师	李文平
课程名称	语言科学与跨学科素养		
论文题目	《财会月刊》刊文计量研究 ——与普通类、理论类、国贸类文本对比		

一、参赛范围

2023年春季学期至2023年秋季学期，全国高校学生修读通识课程撰写的课程论文，或在讨论稿、读书报告等基础上修改而成的论文。。

二、写作要求

(一) 明确的问题意识，清晰的论证思路，充分的征引（文献、数据等材料）；

(二) 文笔流畅，逻辑严密，结论明确；

(三) 体现通识课程的能力诉求：反思能力、学术视野、贯通能力、学术想象力、学术表述能力等；能够体现对重要议题的分析和论证，对关键文本义理的解释和阐发。推荐进行跨学科的写作。

(四) 正文字数一般在5000字以内（至多不超过8000字），须使用中文写作。课程论文应包括论文题目、摘要、关键词、参考文献等部分。如有引文，须注明出处。文章按word格式A4纸（“页面设置”按word默认值）编排，提交电子版。字体：宋体；字号：小四号；字符间距：标准；行距：20磅。具体格式参加附件2。

《财会月刊》刊文计量研究
——与普通类、理论类、国贸类文本对比

【摘要】

近年来，随着中国特色管理科学建设进程的不断推进，《财会月刊》刊文研究成果丰硕。源于既有的学者研究思路，采用文献计量学、对比研究和相关性分析的方法，探究《财会月刊》和普通类、理论类、国贸类文本的词频结构，并对相关性分析筛选后关键词作对比研究；进行语料数据语言网络构建与语言网络分析的探索尝试，从语言网络视角辅助分析。最终得出《财会月刊》有关热点主题、反映先进理论和研究倾向方面的结论，并总结经验反思不足，明确未来进一步探究切实可行的思维指向。

【关键词】

财会月刊 文献计量学 对比研究 语料分析 数据可视化

【正文】

一、研究背景

（一）研究背景概述

党的十八大以来，习近平总书记高度重视中国话语体系建设，强调要用中国理论解释中国实践，用中国实践升华中国理论，加快构建中国特色哲学社会科学。建设中国特色哲学社会科学，离不开中国特色管理科学的贡献。

随着中国特色管理科学的发展，各领域重点期刊的刊文计量研究广受关注，有关学术热点问题的研究势头正盛。刊文计量研究以语言学为工具，各领域刊文为研究对象，将文献计量学与刊文代表的学科相结合，跨学科研究学术问题，研究成果丰硕。

《财会月刊》作为我国地位举足轻重的财会刊物，具有丰富的语料研究价值。关于《财会月刊》的刊文研究，利于从语言学角度理解我国财会领域理论和实务发展成果，为广大财会理论和实务工作者更好地把握《财会月刊》载文的热点主题提供参考和借鉴。

（二）语料数据来源和数据量

本文涉及的语料数据来源有四，分别是：《财会月刊》、《人民日报》、人民日报理论版、《国际贸易》。其中，《人民日报》代表普通类语料，人民日报理论版代表理论类语料，《国际贸易》代表国贸类语料。《人民日报》和人民日报理论版语料数据来自人民日报图文数据库，《财会月刊》和《国际贸易》语料数据来自中国知网学术期刊数据库。

《财会月刊》作为我国重要的财会刊物，多次被评为全国优秀经济期刊和全国中文核心期刊，具有非常广泛的学术影响力，为实务界和理论界共同信赖，可以作为财会类文献计量学研究材料。

《人民日报》于 1946 年创刊并持续发行至今,具备高度的权威性与广泛的影响力,作为现代汉语新闻文体的典型代表,可以为语言研究提供长期稳定的文本材料。

在《人民日报》各版面中,人民日报理论版是人民日报最具重量的版面,是人民日报的优势版面,也是当前学者,党政工作人员的谏言献策窗口。该版面密切关注思想动态、理论前沿、舆论热点,积极宣传阐释党的理论、路线、方略,深入宣传党中央重大决策部署和党的最新理论成果,具有时代色彩鲜明、思想视野宏阔、立足社会现实、主题重要重大等特点。该版面可以为先进理论语料研究提供丰富材料。

《国际贸易》创刊于 1982 年,是由中华人民共和国商务部主管,中国商务出版社主办的国际经济贸易专业期刊,具有广泛的学术影响力,可以作为可靠的语料对比研究材料。

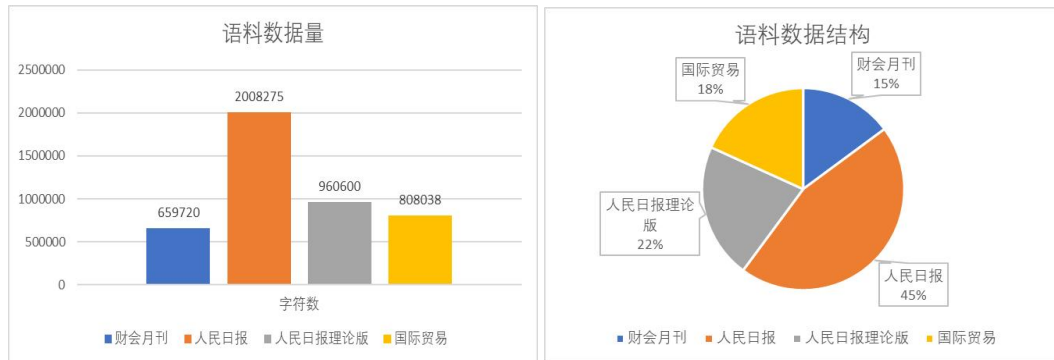


图 1 图 2 语料数据量和数据结构可视化

从语料数据量看,《财会月刊》导入 2023 年 16 期到 22 期的所有刊文内容,以字符为单位,数据量级 66 万。《人民日报》导入 2023 年 10 月 1 日至 12 月 9 日共 70 天的内容,以字符为单位,数据量级 200 万,其中剔除产经、财经、消费、经济、广告版面;剔除广告、图片报道、致读者报眼;剔除地点信息内容(剔除原因和依据见“四、结论思考”部分)。人民日报理论版导入 2023 年全年的所有内容,以字符为单位,数据量级 96 万。《国际贸易》导入 2015 年至 2023 年期刊语料数据,以字符为单位,数据量级 80 万。

从语料数据结构看,《财会月刊》和《国际贸易》占比相近,人民日报理论版占比高于《财会月刊》,《人民日报》占比接近总语料数据的一半。此数据结构和本文将《财会月刊》与国贸类、普通类、理论类相对比的研究思路相契合,具体会在“四、结论思考”部分作出分析。

二、研究目的

利用 Emeditor、Excel 宏代码和 Python 进行数据收集与处理,通过文献计量和对比研究方法,深挖语料数据,对比多类文本,实现对《财会月刊》热点主题和《财会月刊》对于先进理论反映程度的分析。在此基础上,进行语料

数据的语言网络构建与语言网络分析的探索尝试，从语言网络视角辅助分析探究。通过实践检验课堂学习成果，用探究形式反映个人素养提升效果，展示探究拓展能力。

三、过程记录

由于本报告四类语料数据的收集、处理过程几乎无区别，此部分以语料数据量最大的《人民日报》语料数据收集处理过程为主，其余三类语料数据收集处理过程为辅，展示全过程记录。

(一) 数据收集

1. 原始语料数据采集

进入人民日报图文数据库，在所研究时间范围内，以人工筛选方式剔除产经、财经、消费、经济、广告版面，剔除广告、图片报道、致读者报眼，将符合要求的所有文本数据导入 Emeditor。

进入中国知网学术期刊数据库，将研究范围内各期刊文，以 PDF 格式下载，利用 Python 代码将 PDF 格式转为 TXT 格式，导入 Emeditor。



图 3 人民日报图文数据库网页截图

```
import os
import fitz
from pdf2image import convert_from_path

pytesseract.pytesseract_tesseract_cmd = r'c:\Users\何英\appdata\local\programs\python\python38\lib\site-packages\tesseract.exe'

def pdf_to_txt(folder_path):
    # 遍历文件夹中的所有文件
    for root, _, files in os.walk(folder_path):
        for file in files:
            # 检查文件是否为PDF
            if file.endswith('.pdf'):
                # 获取文件完整路径
                pdf_path = os.path.join(root, file)
                # 将PDF转换为图片
                images = convert_from_path(pdf_path)
                # 遍历输出文件路径
                output_path = os.path.splitext(pdf_path)[0] + '.txt'
                # 将图片转换为TXT
                images_to_txt(images, output_path)
                print(f'{pdf_path} 转换为 {output_path}')

if __name__ == '__main__':
    folder_path = 'D:\Python\语言科学与跨学科案例\期末报告_初始语料' # 修改为您的文件夹路径
    pdf_to_txt(folder_path)
```

图 4 PDF 格式转 TXT 格式代码

2. 分词和数据预处理

首先利用 Emeditor 查找、替换功能，应用正则表达式将文本转换为一句一行形式。

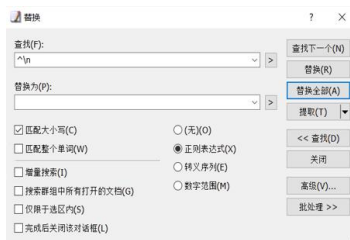


图 5 图 6 文本一句一行形式和运用的正则表达式

使用 jieba 分词，将一句一行文本转换为分词形式的文本。对分词后的文本应用正则表达式，删去空白行、标点、英文字母，按词性筛选出名词、动词和形容词，实现数据预处理。

```
# -*- coding: utf-8 -*-
#2023-9-21使用jieba分词
#copyright @liyifan

import jieba
import jieba.posseg as pseg

datafile = open('D:\人民日报语料收集_02_一句一行.txt', encoding = 'utf-8')
words = pseg.cut(datafile.read())

for word in words:
    print(str(word))
```

图 7 分词代码

- 1 中央+
- 2 金融+
- 3 工作+
- 4 会议+
- 5 在+
- 6 北京+
- 7 举行+
- 8 习近平+
- 9 李强+
- 10 作+
- 11 重要讲话+
- 12 赵乐际+
- 13 王沪宁+
- 14 蔡奇+
- 15 薛祥+
- 16 李希+
- 17 出席+
- 18 金融+
- 19 是+
- 20 国民经济+
- 21 的+
- 22 金融+
- 23 是+
- 24 国家+

图 8 分词和数据预处理后的文本

(二) 数据处理、改进与探索

1. 制作词云图和频度表

使用 Python 代码，将分词和数据预处理的结果制成词云图。

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# 从文件中读取词库数据
word_freq = {}
with open('D:\人民日报理论版面\词云图语料.txt', 'r', encoding='utf-8') as file:
    for line in file:
        parts = line.strip().split()
        if len(parts) == 2:
            word, freq = parts
            word_freq[word] = int(freq)

# 创建wordCloud对象
wordcloud = WordCloud(
    width=1800,
    height=900,
    background_color='white',
    colormap='viridis', # 使用更鲜艳的颜色映射
    font_path='C:\WINDOWS\FONTS\SIMGAL.TTF',
    prefer_horizontal=1 # Set prefer_horizontal to 1 to prevent word rotation
)

# 生成词云图
wordcloud.generate_from_frequencies(word_freq)

# 使用Matplotlib显示词云图
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off") # 不显示坐标轴
plt.show()
```

图 9 词云图代码



图 10 词云图一览（从左上到右下依次为《财会月刊》《人民日报》《国际贸易》和人民日报理论版）

将分词和数据预处理后的语料导入 Excel，利用简单的 if 函数，结合筛选和排序功能，制成频度表。统计得出单词、词频和词频占比数据。

1	freq	
20	阿	19 x
26	阿卜杜拉	6 x
29	阿卜杜勒	3 x
38	阿布	9 x
46	阿布扎比	8 x
48	阿尔舒	2 x
57	阿方	9 x
60	阿富汗	3 x
72	阿根廷	12 x
74	阿公	2 x
76	阿霍	2 x
82	阿克苏	6 x
94	阿克苏地	12 x
100	阿克苏河	6 x
113	阿拉伯	13 x
115	阿拉伯半	2 x
117	阿拉伯联	2 x
120	阿拉木图	3 x
124	阿里	4 x
136	阿联酋	12 x
139	阿塞拜疆	3 x
145	阿斯坦	6 x
147	阿罐西	2 x
153	阿依河	6 x
155	阿姨	2 x
160	阿中	5 x
188	啊	28 x
190	哀悼	2 x
195	埃	5 x

1	freq	
2	的	17531
3	和	4609
4	在	4606
5	发展	3966
6	了	3862
7	是	3067
8	中国	2375
9	为	2363
10	新	2280
11	等	2140
12	建设	1542
13	与	1535
14	上	1417
15	要	1356
16	文化	1341
17	中	1309
18	工作	1298
19	推动	1257
20	经济	1237
21	年	1214
22	合作	1176
23	月	1160
24	对	1158
25	到	1097
26	以	1093
27	有	1066
28	习近平	1059
29	创新	1048
30	我们	1041

1	word	freq	
5	发展	3966	
8	中国	2375	
12	建设	1542	
16	文化	1341	
18	工作	1298	
19	推动	1257	
20	经济	1237	
22	合作	1176	
28	习近平	1059	
29	创新	1048	
34	服务	944	
35	国家	939	
36	推进	930	
37	重要	899	
40	企业	889	
41	好	857	
43	多	851	
46	坚持	819	
47	国际	809	
49	提升	796	
51	大	763	
53	持续	749	
54	金融	748	
55	时代	732	
56	产业	728	
57	实现	707	
58	加强	690	
61	思想	649	
62	世界	647	

1	word	freq	词频占比
2	发展	3966	2.929488
3	中国	2375	1.754295
4	建设	1542	1.138999
5	文化	1341	0.99053
6	工作	1298	0.958769
7	推动	1257	0.928484
8	经济	1237	0.913711
9	合作	1176	0.868653
10	习近平	1059	0.782231
11	创新	1048	0.774106
12	服务	944	0.697286
13	国家	939	0.693593
14	推进	930	0.686945
15	重要	899	0.664047
16	企业	889	0.65666
17	好	857	0.633024
18	多	851	0.628592
19	坚持	819	0.604955
20	国际	809	0.597568
21	提升	796	0.587966
22	大	763	0.56359
23	持续	749	0.553249
24	金融	748	0.552511
25	时代	732	0.540692
26	产业	728	0.537738
27	实现	707	0.522226
28	加强	690	0.509669
29	思想	649	0.479384
30	世界	647	0.477907

图 11 频度表形成过程（以《人民日报》文本为例）

2. 数据改进

利用 Excel 宏代码，将《财会月刊》频度表数据和《人民日报》、人民日报理论版、《国际贸易》对比，自动计算统计单词的词频占比差值，利用相关性分析的原理，在两语料数据词频占比大于等于 0.2%的词中，将词频占比差值大于等于 1%的词作为关键词筛选出来，进行数据对比分析。

但在实际操作中，由于文本数据量过大，Excel 宏代码运行速度过慢，处理过程卡顿，并没有使用 Excel 宏代码完成相关性分析的操作，取而代之的是更为高效的 Python 代码。

```

(通用) CompareWordFrequency
Sub CompareWordFrequency()
    Dim ws1 As Worksheet, ws2 As Worksheet
    Dim lastRow1 As Long, lastRow2 As Long, i As Long, j As Long
    Dim word1 As String, frequency1 As Double, word2 As String, frequency2 As Double
    Dim resultColumn As Integer

    Set ws1 = ThisWorkbook.Worksheets("Sheet1")
    Set ws2 = ThisWorkbook.Worksheets("Sheet2")

    lastRow1 = ws1.Cells(ws1.Rows.Count, "A").End(xlUp).Row
    lastRow2 = ws2.Cells(ws2.Rows.Count, "A").End(xlUp).Row

    resultColumn = lastRow1 + 1

    For i = 2 To lastRow1
        word1 = ws1.Cells(i, 1).Value
        frequency1 = ws1.Cells(i, 2).Value
        For j = 2 To lastRow2
            word2 = ws2.Cells(j, 1).Value
            frequency2 = ws2.Cells(j, 2).Value
            If word1 = word2 And Abs(frequency1 - frequency2) >= 1 Then
                ws1.Cells(i, resultColumn).Value = word1 & " " & frequency1 & " vs " & word2 & " " & frequency2
            End If
        Next j
    Next i
End Sub

```

图 12 对比分析宏代码示例

```

import pandas as pd

# 读取两个Excel文件
df1 = pd.read_excel('财会 对比.xlsx')
df2 = pd.read_excel('人民日报 对比.xlsx')

# 创建一个空的DataFrame用于存储结果
result = pd.DataFrame(columns=['词', '词频占比1', '词频占比2'])

# 遍历第一组数据
for index, row in df1.iterrows():
    word = row['词']
    freq_ratio1 = row['词频占比']

    # 在第二组数据中查找相同的词
    matched_row = df2[df2['词'] == word]

    # 如果找到了相同的词，并且词频占比相差大于等于1
    if not matched_row.empty and abs(matched_row['词频占比'].values[0] - freq_ratio1) >= 1:
        result = result._append({'词': word, '词频占比1': freq_ratio1, '词频占比2': matched_row['词频占比'].values[0]}, ignore_index=True)

# 将结果保存到新的Excel文件中
result.to_excel('result对比.xlsx', index=False)

```

图 13 对比分析 Python 代码

词	词频占比1	词频占比2	词频占比差值
企业	6.5047345	0.65666041	5.848074046
研究	3.4376287	0.26886883	3.168759824
发展	2.6142445	5.05248675	-2.438242206
经济	2.0172911	0.91371083	1.103580233
治理	1.399753	0.30875596	1.09099702
环境	1.3585838	0.20903813	1.14954565
政策	1.3585838	0.25114122	1.107442564
风险	1.2762454	0.23119765	1.04504772
中国	0.3911075	1.75429525	-1.363187801

表 1 《财会月刊》与普通类文本关键词对比结果

3. 尝试构建语言网络

制作出词云图、频度表和对比数据后，进一步尝试探索制作语言网络。最终利用 Python 和频度表数据，制作出了简单语言网络。但是这一简单语言网络并不属于真正的语义网络，只是对于语言网络构建的尝试，可以辅助数据分析与结论得出，并不能作为主要的分析依据。

```

import networkx as nx
import matplotlib.pyplot as plt
from matplotlib import font_manager

with open('D:\期末报告\语言网络.txt', 'r', encoding='utf-8') as file:
    lines = file.readlines()

words = []
frequencies = []
for line in lines:
    word, frequency = line.strip().split()
    words.append(word)
    frequencies.append(int(frequency))

G = nx.Graph()

for i in range(len(words) - 1):
    G.add_node(words[i])
    G.add_node(words[i+1])
    G.add_edge(words[i], words[i+1])

pos = nx.spring_layout(G)

font_path = None
font_prop = font_manager.FontProperties(fname=font_path)
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False

nx.draw(G, pos, with_labels=True, node_color='lightblue', edge_color='gray')
plt.show()

```

图 13 语言网络代码

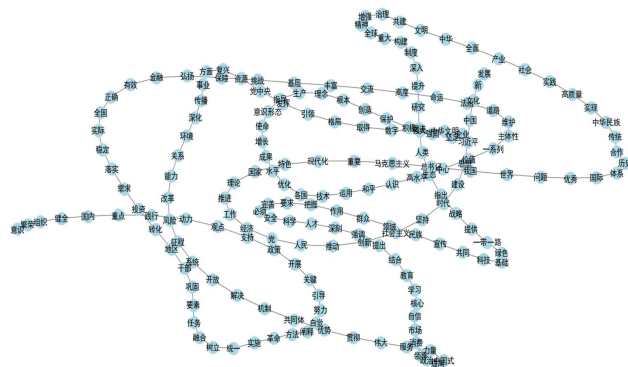


图 14 语言网络尝试（以人民日报理论版为基础）

由图 13 所示代码制成的图 14 语言网络并没有实际的研究价值，因为此语言网络既不能反映词频也不能反映词之间的联系。在此基础上，对语言网络代码进行进一步改进（图 16），最终得到图 15 的语言网络，此图蓝色节点的大小与该节点表示单词词频正相关，边的疏密程度与语料复杂程度正相关。该语言网络可以作为辅助数据结果，参与研究问题的分析。但受限于时间、资源，此

语言网络与真正的语义网络仍有差距。此语言网络边的绘制是根据节点数量和 random 函数随机生成的，其疏密程度只能一定程度上反应语料的复杂程度，并不能反映各节点所代表的词义的真正联系。

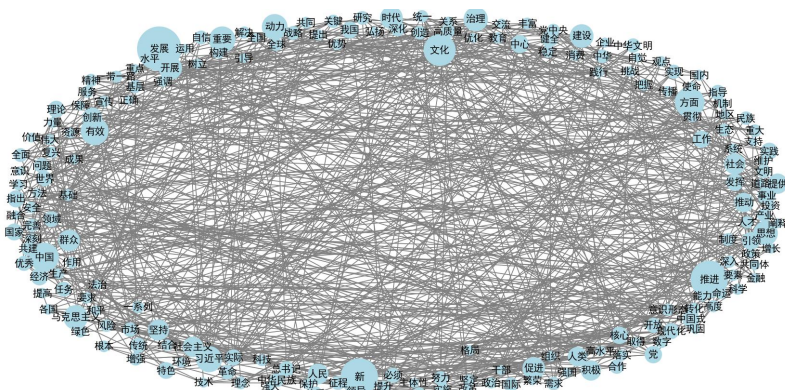


图 15 语言网络尝试改进（以人民日报理论版为基础）

```
import networkx as nx
import matplotlib.pyplot as plt
import random
from matplotlib import font_manager

with open('D:\人民日报理论版面\语料.txt', 'r', encoding='utf-8') as file:
    lines = file.readlines()

words = []
frequencies = []
for line in lines:
    word, frequency = line.strip().split()
    words.append(word)
    frequencies.append(int(frequency))

G = nx.Graph()

for i in range(len(words) - 1):
    word1 = words[i]
    word2 = words[i + 1]
    freq1 = frequencies[i]
    freq2 = frequencies[i + 1]

    G.add_node(word1, freq=freq1)
    G.add_node(word2, freq=freq2)

    G.add_edge(word1, word2, weight=freq1)
    G.add_edge(word2, word1, weight=freq2)

    extra_edges = random.randint(1, 2)
    for _ in range(extra_edges):
        other_word = random.choice([w for w in words if w != word1 and w != word2])
        G.add_node(other_word, freq=random.randint(1, frequencies.index(max(frequencies)) + 1))
        G.add_edge(other_word, word1, weight=random.randint(1, freq1))
        G.add_edge(other_word, word2, weight=random.randint(1, freq2))

font_path = None
font_prop = font_manager.FontProperties(fname=font_path)
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False

pos = nx.spring_layout(G, k=100)
node_size = [freq * 1 for freq in frequencies]
nx.draw(G, pos, with_labels=True, node_color='lightblue', node_size=node_size, edge_color='gray')
plt.show()
```

图 16 改进后的语言网络代码（以人民日报理论版为基础）

四、结论思考

（一）数据处理结果与分析

1. 《财会月刊》语料数据分析

《财会月刊》研究的主体集中在“企业”，围绕此主体，产生的主题有业财一体化与新兴技术（“数字化”“信息”“转型”“创新”“ChatGPT”）、审计（“审计”“成本”“审计师”）、宏观环境研究（“环境”“经济”“政策”“政府”）、生态与经济（“生态”“金融”“经济”“环境”“绿色”“A股”“ESG”）、会计业务（“财务”“绩效”“董事”“资产”“成本”“审计”）等等，主题范畴广，内容多，涉及财会领域各大重点热点问题。

前 20 词中，“审计”“转型”居于前列，反映出我国经济从高速发展转向高

质量发展对财会领域的影响，尤其是审计重要性和转型意识的提升；“数字化”“信息”“创新”均居于前 10 位，反映出业财一体化的热点话题；“政策”“政府”“治理”反映出对政府管理的学术反馈；“碳”居于第 20 位，反映出双碳目标下财会领域相关问题的理论和实务研究势头正盛。

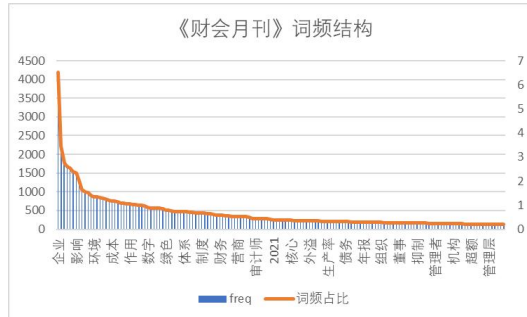


图 17 《财会月刊》词频结构

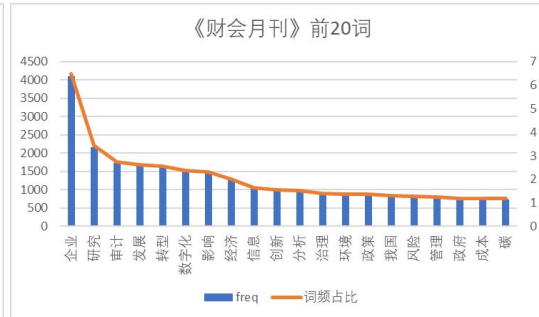


图 18 《财会月刊》前 20 词

单词	词频	词频占比	单词	词频	词频占比
企业	4108	6.504734	分析	962	1.523261
研究	2171	3.437629	治理	884	1.399753
审计	1729	2.737752	环境	858	1.358584
发展	1651	2.614245	政策	858	1.358584
转型	1612	2.552491	我国	832	1.317415
数字化	1508	2.387814	风险	806	1.276245
影响	1469	2.32606	管理	793	1.255661
经济	1274	2.017291	政府	754	1.193907
信息	1027	1.626184	成本	741	1.173322
创新	975	1.543845	碳	741	1.173322

表 2 《财会月刊》前 20 词

2. 《财会月刊》与普通类文本：《人民日报》的对比分析

从词频结构看，普通类文本《人民日报》的词频结构与《财会月刊》最大的区别在于，前者词频占比总体低于后者，前者单词种类数量远大于后者，这体现出报纸和学术期刊在语言丰富程度上的差别。从前 20 词看，《财会月刊》聚焦专业领域主题研究，以企业为突出主体，《人民日报》涉及主体主题广泛全面，突出中国特色。

从文本关键词词频占比对比分析看，《财会月刊》突出“企业”的主体地位和“研究”的学术性质，更多倾向于经济研究、治理问题、政策解读、风险解析等主题；《人民日报》突出“中国”主体，倾向于“发展”相关的主题。

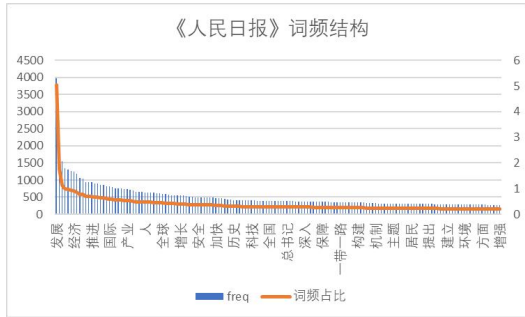


图 19 《人民日报》词频结构

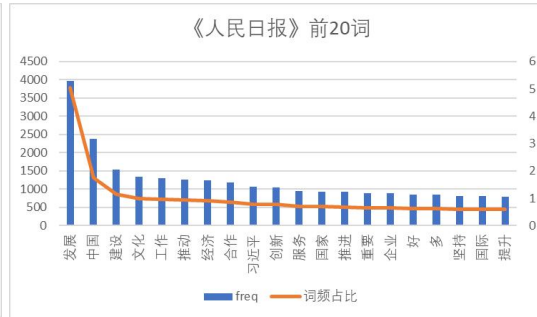


图 20 《人民日报》前 20 词

词	词频占比1	词频占比2	词频占比差值
企业	6.504734459	0.656660413	5.848074046
研究	3.437628654	0.26886883	3.168759824
发展	2.614244545	5.052486751	-2.438242206
经济	2.017291066	0.913710833	1.103580233
治理	1.399752985	0.308755965	1.09099702
环境	1.358583779	0.209038129	1.14954565
政策	1.358583779	0.251141215	1.107442564
风险	1.276245368	0.231197648	1.04504772
中国	0.391107452	1.754295253	-1.363187801

表 3 《财会月刊》与《人民日报》文本关键词词频占比对比结果

3. 《财会月刊》与理论类文本：人民日报理论版的对比分析

从词频结构看，人民日报理论版和《人民日报》基本一致，均表现出词频占比平均低于《财会月刊》，词频种类数量远远大于《财会月刊》的特点。从前 20 词看，人民日报理论版突出“文化”“理论”“思想”等理论类特征性主题。

从文本关键词词频占比对比分析看，和普通类相比，人民日报理论版与《财会月刊》在关键词“企业”“中国”上的词频占比差值增大，表现出《财会月刊》的学术性质和人民日报理论版具有时代色彩鲜明、思想视野宏阔、立足社会现实、主题重要重大等特点。和普通类结果相比，无论是从词频占比差异较大词的数量角度，还是从词频占比差值角度，《财会月刊》与理论类的差异均显著减少，证明《财会月刊》紧跟理论步伐，将理论思想转化为财会领域学术理论与实务。

辅助以语言网络来分析，从语言网络各节点边的疏密程度看，人民日报理论版明显密集于《财会月刊》，说明前者语言比后者丰富。

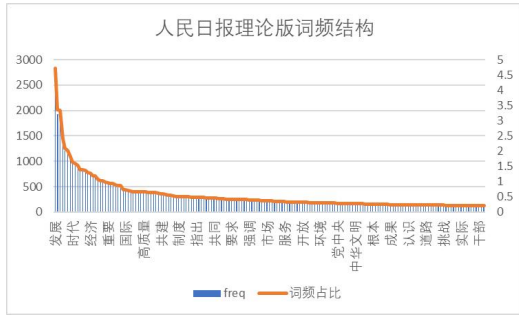


图 21 人民日报理论版词频结构



图 22 人民日报理论版前 20 词

词	词频占比1	词频占比2	词频占比差值
企业	6.504734459	0.271327942	6.233406517
研究	3.437628654	0.499173841	2.938454813
发展	2.614244545	4.715192625	-2.10094808
环境	1.358583779	0.299156448	1.059427331
中国	0.391107452	2.414122967	-2.02301552

表 4 《财会月刊》与人民日报理论版文本关键词词频占比对比结果

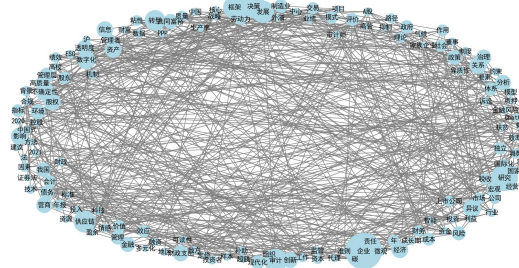


图 23 《财会月刊》语言网络

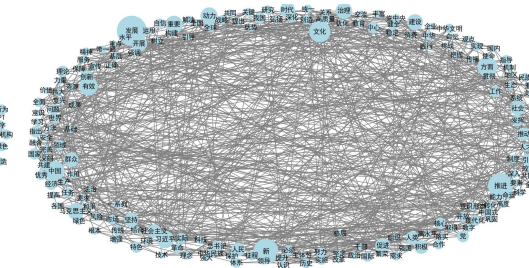


图 24 人民日报理论版语言网络

4. 《财会月刊》与国贸类文本：《国际贸易》的对比分析

从词频结构看，《国际贸易》单词种类数和平均词频占比均低于《财会月刊》，不过这与《国际贸易》每期刊文数量少、期间间隔长有关。从前 20 词看，《国际贸易》更多倾向于国际贸易、中美、WTO 等话题，和《财会月刊》有明显差异，表现专业学术性质。

从文本关键词词频占比对比分析看，二者在“企业”一词的差异较小，在“投资”“中国”的差异较大，反映出两期刊研究倾向的显著差异。但二者词频占比有明显差异的关键词数量较少，体现出两期刊都属于商科范畴，语言和研究对象趋于一致。

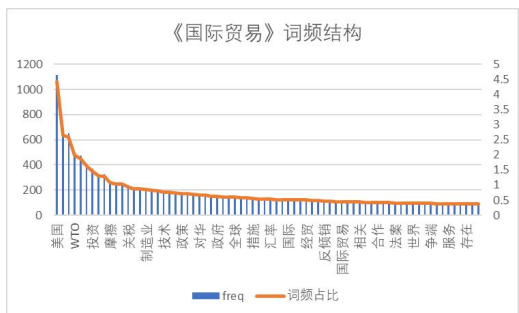


图 25 《国际贸易》词频结构

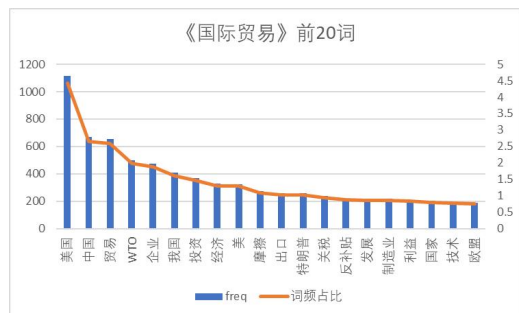


图 26 《国际贸易》前 20 词

词	词频占比1	词频占比2	词频占比差值
企业	6.504734459	1.87524831	4.629486148
发展	2.614244545	0.86213747	1.75210708
影响	2.326060107	0.38140644	1.944653671
创新	1.543845204	0.38140644	1.162438768
中国	0.391107452	2.64998014	-2.258872683
投资	0.205846027	1.47000397	-1.264157946

表5 《财会月刊》与《国际贸易》文本关键词词频占比对比结果

(二) 结论

《财会月刊》侧重于以企业为主体，财会领域当下突出热点问题研究，专业性和学术性强，兼顾理论与实务。《财会月刊》充分应用和反馈党中央的先进思想理论，将先进思想深入财会，使先进理论落实落地，推动中国特色管理科学发展。

(三) 总结与反思

由于学期时间和知识积累有限，本报告并不能更加深入专业地理解认识所探究问题，为了一定程度上弥补深度不足的缺陷，本报告收集了大量语料数据，并应用多种研究方法，多角度探究问题，力求构建完整地逻辑链条，科学地实现分析探究。

细节操作上，将课堂和二课学习交流的成果运用于问题探究过程中，用Excel的宏代码和函数结合Emeditor替换功能实现大量级数据的处理，用Python解决全流程各种复杂问题。但也存在受时间、资源限制，没能制得真正意义上语义网络的问题。

在此报告基础上，未来可以通过继续探究语义网络的构建，深挖各类语料数据尤其是《财会月刊》的潜在研究价值。

最后由衷地感谢先后由李文平老师、李金满老师、孟雅萱老师、李健老师和郭鸿杰老师带来的精彩而具有启发意义的课堂，特别要感谢李文平老师对我个人的启发指导——充实有趣的课堂和广泛拓展的二课指引我走出了研究尝试和技术学习的第一步！学无止境，未来我会继续努力！

参考文献

- [1]单佳美,龙海军. 我国本土会计学理论体系构建的实践与探索——基于北大中文核心期刊会计学类期刊刊文的分析 [J]. 商业会计, 2023, (14): 110-115.
- [2]陈晓芳, 蒋武. 《财会月刊》刊文计量研究: 核心作者、主要机构与热点主题[J]. 财会月刊, 2020, (01): 86-91.
- [3]梁修明,李军涛. 融媒时代理论传播何以“破壁出圈”——以人民日报理论版“思想纵横”栏目为例 [J]. 中国报业, 2022, (01): 126-128. DOI:10.13854/j.cnki.cni.2022.01.058
- [4]王永贵,汪寿阳,吴照云等.深入贯彻落实习近平总书记在哲学社会科学工作座谈会上的重要讲话精神加快构建中国特色管理学体系[J].管理世界,2021,37(06):1-35.DOI:10.19744/j.cnki.11-1235/f.2021.0075