

# 基于文本挖掘对新旧保险合同准则的分析

韩莉婷<sup>1</sup>, 阎虎勤<sup>2</sup>

<sup>1</sup>厦门大学嘉庚学院会计与金融学院, 福建 漳州

<sup>2</sup>厦门国家会计学院信息管理处, 福建 厦门

收稿日期: 2023年6月18日; 录用日期: 2023年7月18日; 发布日期: 2023年7月27日

## 摘要

新保险合同准则于今年的1月1日开始施行, 它标志着我国进一步完善企业会计准则体系, 保持与国际财务报告准则的持续趋同。因此, 本文利用Python程序语言, 通过网页爬虫对2020年发布的新旧保险合同准则进行文本挖掘。另外下载我国财政部于2006年和2009年印发的三份旧保险合同准则, 通过本地文本导入, 利用jieba库进行文本分词, collection库统计词频, wordcloud库制作词云等, 进行对比、联系, 发现新旧保险合同都出现如“合同”、“计量”等词汇。总结出新保险准则着重关注在保险合同定义、保险合同计量模型等方面。同时梳理得出新保险合同准则的发布实施有利于凸显“保险姓保”的原则; 有利于抑制保险合同美化财务业绩; 还有利于提升我国保险行业的国际形象。

## 关键词

高频词, 计量, 保险公司

# Analysis of Old and New Insurance Contract Guidelines Based on Text Mining

Liting Han<sup>1</sup>, Huqin Yan<sup>2</sup>

<sup>1</sup>Accounting and Finance, Xiamen University Tan Kah Kee College, Zhangzhou Fujian

<sup>2</sup>Information Management Division, Xiamen National Accounting Institute, Xiamen Fujian

Received: Jun. 18<sup>th</sup>, 2023; accepted: Jul. 18<sup>th</sup>, 2023; published: Jul. 27<sup>th</sup>, 2023

## Abstract

The new insurance contract standard came into effect on January 1 this year, marking the further improvement of China's accounting standards system for enterprises and maintaining continuous convergence with IFRS. Therefore, this paper uses the Python programming language to conduct

text mining of the old and new insurance contract guidelines released in 2020 through web crawlers. In addition, download the three old insurance contract guidelines issued by China's Ministry of Finance in 2006 and 2009, import through local text, use the jieba library for text word segmentation, collection library to count word frequency, wordcloud library to make word clouds, etc. For comparison and contact, and found that new and old insurance contracts appear words such as "contract" and "measurement". The new insurance standard focuses on the definition of insurance contracts and the measurement model of insurance contracts. At the same time, it is concluded that the issuance and implementation of the new insurance contract standards is conducive to highlighting the principle of "insurance surname protection". It is conducive to suppressing insurance contracts and beautifying financial performance; it is also conducive to enhancing the international image of China's insurance industry.

## Keywords

High Frequency, Word Metering, Insurance Company

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着我国经济的高质量发展和物质产品的极大丰富,人们对平抑财产风险和对抗生老疾病风险的意识逐渐增强,保险业迎来了黄金发展机遇期[1]。2020年12月,财政部修订发布了《企业会计准则第25号——保险合同》,以下简称“新保险合同准则”。不再执行2006年2月印发的《财政部关于印发〈企业会计准则第1号——存货〉等38项具体准则的通知》(财会〔2006〕3号)中的《企业会计准则第25号——原保险合同》和《企业会计准则第26号——再保险合同》,以及财政部于2009年12月印发的《保险合同相关会计处理规定》(财会〔2009〕15号),以下简称“旧保险合同准则”[2]。大数据时代下数据总量大幅度增加,传统数据分析和处理技术存在很大限制,为了探究新旧保险合同准则的差异,本文结合python程序语言爬取新旧保险合同准则进行文本分析,通过大数据挖掘,提供了可视化的数据分析结果,梳理出新旧保险合同准则中的异同点。

## 2. 文献综述

随着2020年新保险合同准则的发布,学界加大了对保险合同准则领域的关注。梁建民(2020)认为新保险合同准则实施对保险行业财务管理模式产生了重大影响[1]。杨征(2021)认为新保险合同准则揭示了寿险公司的经营实质,更是引导了行业关注资产负债匹配与资产管理,以实现企业高质量发展[3]。李华(2023)对新准则与旧准则的信息披露要求进行分析,认为新准则下准备金计量、信息披露更加透明、丰富[4]。

国际会计准则理事会(IASB)2017年发布了《国际财务报告准则第17号——保险合同》(IFRS 17)。经济全球化发展的当下,我国的保险合同准则逐步向国际准则趋同。许闲(2019)分析IFRS 17关于保险合同准则确认、计量与披露重大调整,讨论IFRS 17和IFRS 9与我国当前会计实务的差异与影响[5]。孙火秀、曾诚和王煦楠(2020)认为IFRS 17给保险公司收入带来影响[6]。顾丽雅(2022)分析IFRS 17的变化并明确新准则对国内保险公司的经营模式等方面的影响[7]。

目前学界对于保险合同领域的研究主要集中于相关准则实施后产生的影响, 对新保险合同与旧保险合同准则的差异性关注较少。我国不断完善新准则, 取代旧准则, 其中的更新变化反映了我国积极推进保险行业的发展, 正逐步向国际准则靠拢的决心。另外, 应用 Python 文本挖掘探究保险合同准则异同点的相关文献门可罗雀。

### 3. 数据描述与方法

#### 3.1. 使用数据获取和数据预处理

本文一共采集了新旧两版保险合同准则, 新保险合同准则来源于财政部会计准则委员会官方网站 <https://www.casc.org.cn/2018/0815/213104.shtml>, 其余三份旧保险合同准则来源于财政部会计准则委员会官方网站准则发布中附件的下载。

本文主要通过 Python 各个库函数的运用实现文本的抓取、处理与分析。首先通过 BeautifulSoup4 库函数对 HTML 的解析功能, 将网页上的信息转化成文本信息, 并且通过文本挖掘来处理实现对新保险合同准则的爬取。再运用 jieba 库函数对收集到的文本进行分词处理, 即根据语义将一句话分成几个单词, 例如, “根据本准则作出的相关会计估计处理结果” 分词成: “根据” “本” “准则” “作出” “的” “相关” “会计” “估计” “处理” “结果”。句子里一些毫无意义的停词, 比如文本中出现的 “的” 将其过滤掉, 该过程使用 jieba 自带的停词词库, 实现停词过滤。分词这一步尤为重要, 分词效果的好坏直接影响了后续分析的准确性。依据 jieba 包里自带的词典, 文本被分为一个个的词后。为了提取有用的信息, 需要进一步过滤掉文本中的 “本” 等实词。最后统计新旧保险合同准则的词频, 截取排序前 25 的词, 见表 1。

Collection 包中提供了数据类型和容器, 较好地实现了数据的处理和操作。如 “counter” 用于统计对象中每个元素的出现次数; 使用 “defaultdict” 可以方便地处理字典中不存在的键。为了让统计结果更加直观, 便于寻找主要属性特征, 本文使用 Python 的 Wordcloud 库函数根据新旧保险合同准则制作词云图。Wordcloud 词云方法将文本中不同的词汇按照出现的词频数量进行统计, 如图 1 和图 2 所示。词云以单词为基本单位, 根据单词出现频率来确定词语的大小, 出现频数越多的单词, 字体越大, 直观地在一张图片中展示文本中的关键词。



Figure 1. Old insurance contract word cloud

图 1. 旧保险合同词云



Figure 2. New insurance contract word cloud  
图 2. 新保险合同词云

Table 1. Word frequency of old and new insurance contract guidelines  
表 1. 新旧保险合同准则的词频

词序	新保险合同		旧保险合同	
	词	频数	词	频数
1	合同	581	合同	148
2	企业	271	准备金	143
3	现金流量	132	责任	75
4	确认	123	分保	44
5	服务	123	当期	41
6	相关	103	风险	38
7	变动	99	计量	36
8	价值	97	寿险	35
9	金额	86	赔款	34
10	当期	81	发生	32
11	分出	78	金额	31
12	准则	77	保费	30
13	资产	77	按照	30
14	责任	73	赔付	29
15	调整	73	费用	28
16	亏损	71	确认	27
17	计入	65	分出	27
18	发生	62	计入	25
19	边缘	62	损益	25
20	负债	61	计算	24
21	采用	61	根据	23
22	进行	60	到期	23
23	规定	59	重大	23
24	下列	58	约定	21
25	部分	58	提取	20
合计		2691		1012

## 3.2. 高频词总结

对于收集的排名前 25 位的高频词进行统计, 新保险合同准则的高频词共计 2691 个, 而旧保险合同准则只有 1012 个。因此, 从高频词的数量上可以看出, 新保险合同准则的内容是旧保险合同准则的几倍。其规范的细致性以及完善性应远超旧保险合同准则。准则的完备性能够降低准则运行出错的可能性, 使得新保险合同准则能够更好地为企业财务计量提供基础。

另外, 基于词汇出现频次的高低, 通过观察与分析, 新旧保险合同准则都出现如“合同”、“计量”等词汇。本文结合对新旧保险合同准则的阅读和理解, 得出新保险合同准则在保险合同定义、保险合同计量模型等方面作了较大修改:

### 3.2.1. 完善保险合同定义

新保险合同准则发布之前, 我国旧保险合同分为原保险合同和再保险合同。它们是相辅相成的, 原保险是再保险的基础, 再保险是由原保险派生的, 支持和促进原保险的发展。当时原保险的主体是保险人与投保人, 而再保险的主体是原保险人和再保险人。因此本文对旧保险合同准则中“保险合同”的定义参考我国财政部于 2006 年发布的企业会计准则第 25 号——《原保险合同》。其对于“保险合同”是这样定义的: 保险合同是指保险人与投保人约定保险权利义务关系, 并承担源于被保险人保险风险的协议。除此之外, 原保险合同中还具体规定了“保险合同”是指保险人向投保人收取保费, 对约定的可能发生的事故因其发生所造成的财产损失承担赔偿责任, 或者当被保险人死亡、伤残、疾病或者达到约定的年龄、期限时承担给付保险金责任的保险合同。

2020 年我国财政部修订印发了新保险合同准则, 其中规定“保险合同”: 是指企业与保单持有人约定, 在特定保险事项对保单持有人产生不利影响时给予其赔偿, 并因此承担源于保单持有人重大保险风险的合同。新保险合同准则完善了保险合同的定义, 更清晰地阐述了保险合同的主体为合同签发人和保单持有人; 并明确了“保险风险”是指从保单持有人转移至合同签发人的除金融风险之外的风险要求。保险合同定义的完善, 表明我国加大对保险行业的重视, 从根本法则中查缺补漏, 保护投保人和被投保人双方的利益。

### 3.2.2. 完善保险合同计量模型

从保险运作角度来看, 购买保单的人愿意支付已知数额的金钱, 让保险公司承担未来不确定性带来的财务后果, 保险公司承担了众多投保人的责任。旧保险合同准则中规定: 保险人应当根据在原保险合同延长期内是否承担赔偿责任, 将原保险合同分为寿险原保险合同和非寿险原保险合同。从投资者角度出发, 大多数的终身寿险合同会导致索赔, 但时间不确定; 非寿险合同会导致索赔, 但不可能事先确定好哪些保险合同会导致索赔。因此, 保单的集合分组降低了损失的波动性, 使得风险单位损失更容易预测。

新保险合同准则不再区分保险业务的类型, 而是围绕负债、收入和利润三大会计要素进行一系列调整。在负债方面, 新合同保险准则下保险合同负债计量包括履约现金流量和合同服务边际两个组成部分。其中, 履约现金流量包含未来流量的估计、货币时间价值及金融风险调整、非金融风险调整三个组成部分。值得一提的是, 新保险合同准则明确保险公司在收到保险费时不能立即确认收入, 需要确认为一项保险合同负债, 之后再提供保险服务期间逐期确认收入。

原保险合同准则中没有明确合同服务边际在初始确认后如何反映未来提供的变化。在实践中绝大多数保险公司利用这一漏洞, 将当期作出的有利精算假设调整确认为当期利润, 甚至还通过滥用, 调整公司未来确认的利润。计量方式是基石, 新保险合同准则下改进合同服务边际对提高保险行业整体透明度起着重要作用。

新保险合同准则要求保险公司对于不具有直接参与分红特征的保险合同组, 资产负债表日合同组的合同服务边际账面价值应当以期初账面价值为基础, 经过当期产生的汇兑差额等项目调整后予以确定。这一条例更真实地反映合同服务边际在后续期间的变化, 有利于降低利润操纵的空间, 为财务报表使用者提供决策的有用信息。此外, 新保险合同准则对具有直接参与分红特征的保险合同组计量方法进行特殊规定, 企业应当按照基础项目公允价值扣除浮动收费的差额, 估计具有直接参与分红特征的保险合同组的履约现金流量。

## 4. 结论

新保险合同准则于今年的 1 月 1 日开始施行, 这一准则的发布实施对保险行业产生如下深远影响:

### 4.1. 有利于凸显“保险姓保”的原则

从保险的本质含义出发, 保险合同是被保人最高效的保险方案。保险公司作为风险转移的关键角色, 在开发保险产品时不能着眼于遇到风险只赔付金额, 而要更加关注如何解决通货膨胀因素带来的财务困境。新保险合同准则的发布, 规范了保险公司的各项经营活动, 促进保险公司根据保障需求来开发效率更高的保险产品。

### 4.2. 有利于抑制保险公司美化财务业绩

保险公司的盈利基础是对于产品风险概率的准确计算, 因而计量方式起着至关重要的作用。原保险合同准则的精算假设调整对未来利润的影响直接计入当期损益, 一些保险公司虚假调整精算假设达到美化财务业绩的目的。新保险合同准则改进了合同服务边际计量的方式, 精算假设调整对未来利润的有利影响不允许计入当期损益, 必须在未来提供服务的期间逐步确认, 使得保险公司利用调整精算假设来调节当期利润的目的落空, 一定程度上抑制了利润操纵行为, 有助于提高会计信息质量[8]。

### 4.3. 有利于提升我国保险行业国际形象

随着经济全球化的深入, 保险业的国际化趋同。国际会计准则理事会分别于 2017 年 5 月和 2020 年 6 月发布《国际财务报告准则第 17 号——保险合同》和《对〈国际财务报告准则第 17 号〉的修订》, 自 2023 年 1 月 1 日起实施。国际财务报告准则的发布使各国保险公司的财务报表更具有可比性。与国际财务报告准则趋同的新保险合同准则代表了当前我国保险会计领域的先进水平。该准则的发布实施, 有利于我国保险公司与国际同行对标, 提升我国保险行业的国际形象和国际影响力。

## 致 谢

本论文是在阎虎勤老师开设的《大数据基础与应用》课程作业论文的基础上完成的, 感谢授课阎虎勤老师, 有幸求学于恩师门下, 不以吾之愚笨, 耐心蒙之。

## 项目基金

本论文得到了厦门市科学技术协会(厦科协[2023]5 号) 2023 年重点调研课题“高校毕业生在厦就业及就业前职业技能培训情况调研”项目和厦门市行为科学学会的支持。

## 参考文献

- [1] 梁建民. 新保险合同会计准则实施对保险企业的影响探析[J]. 齐鲁珠坛, 2022(1): 12-16.
- [2] 中华人民共和国财政部. 关于修订印发《企业会计准则第 25 号——保险合同》的通知[EB/OL].

[https://www.gov.cn/zhengce/zhengceku/2020-12/25/content\\_5573188.htm](https://www.gov.cn/zhengce/zhengceku/2020-12/25/content_5573188.htm), 2020-12-25.

- [3] 杨征. 新保险合同准则揭示险企业经营实质, 引领行业高质量发展[J]. 中国保险, 2021(6): 4.
- [4] 李华. 新保险合同准则下财产保险准备金信息披露及使用探析[J]. 财务与会计, 2023(5): 47-50.
- [5] 许闲. 国际保险会计准则的最新发展及对我国的影响——基于 IFRS 17 和 IFRS 9 [J]. 会计研究, 2019(1): 21-27.
- [6] 孙火秀, 曾诚, 王煦楠. 新保险合同会计准则 IFRS 17 对保险公司收入的影响分析[J]. 保险理论与实践, 2022(1): 82-89.
- [7] 顾丽雅. 新保险合同会计准则(IFRS 17)对保险业的影响初探[J]. 全球流通经济, 2022(28): 157-160.
- [8] 潘艳红. 新保险合同准则对保险公司影响探析[J]. 新会计, 2023(1): 35-39.

---

## 附录

---

### Python 词云图程序代码

---

```
##2020 新保险合同

%matplotlib inline
%config InlineBackend.close_figures = False
import numpy as np
import matplotlib.pyplot as plt

from matplotlib import rcParams
from PIL import Image

import urllib
import bs4
from bs4 import BeautifulSoup

from wordcloud import WordCloud,STOPWORDS
import collections
import jieba
my_url = "https://www.casc.org.cn/2018/0815/213104.shtml"
wp = urllib.request.urlopen(my_url)
soup = BeautifulSoup(wp,'lxml')
text = soup.get_text().strip()
print(text)

text = text.replace(",","")
text = text.replace(";","")
text = text.replace("!","")
text = text.replace("?","")
text = text.replace("&","")
text = text.replace("<","")
text = text.replace(">","")
text = text.replace("(","")
text = text.replace(")", "")
text = text.replace("。","")
text = text.replace(":","")
text = text.replace(":","")
text = text.replace(" ","")
text = text.replace("的","")
text = text.replace("了","")
text = text.replace("在","")
text = text.replace("是","")
text = text.replace("也","")
text = text.replace("\n","")
text = text.replace("\t","")
text = text.replace("\xa0","")
text = text.replace("—","")
text = text.replace(">","")
text = text.replace("\u3000","")
text = text.replace("\r","")
text = text.replace("《","")
text = text.replace("》","")
text = text.replace(",","")
text = text.replace("应当","")
text = text.replace("我们","")
text = text.replace("保险","")
text = text.replace("和","")
```

---



```
text = text.replace("或", "")
text = text.replace("与", "")
text = text.replace("日", "")
text = text.replace("本", "")
text = text.replace("时", "")
text = text.replace("将", "")
text = text.replace("组", "")
text = text.replace("再", "")
text = text.replace("为", "")
text = text.replace("不", "")
text = text.replace("会计准则", "")
```

```
Textsplit = jieba.lcut(text)
print(Textsplit)
```

2018-08-15来源: 会计司

企业会计准则第25号——保险合同

财会[2020]20号

第一章 总则

第一条 为了规范保险合同的确认、计量和相关信息的列报, 根据《企业会计准则——基本准则》, 制定本准则。

第二条 保险合同, 是指企业(合同签发人)与保单持有人约定, 在特定保险事项对保单持有人产生不利影响时给予大保险风险的合同。

保险事项, 是指保险合同所承保的、产生保险风险的不确定未来事项。

保险风险, 是指从保单持有人转移至合同签发人的除金融风险之外的风险。

第三条 本准则适用于下列保险合同:

(一) 企业签发的保险合同(含分入的再保险合同);

(二) 企业分出的再保险合同;

(三) 企业在合同转让或非同一控制下企业合并中取得的上述保险合同。

签发保险合同的企业所签发的具有相机参与分红特征的投资合同适用本准则。

再保险合同, 是指再保险分入人(再保险合同签发人)与再保险分出人约定, 对再保险分出人由对应的保险合同所引

具有相机参与分红特征的投资合同, 是指赋予特定投资者合同权利以收取保证金额和附加金额的金融工具。附加金额相机决定, 且预计构成合同利益的重要部分。

子', '咨询', '委员', '领导', '讲话', '工作', '通知', '准则', '发布', '准则', '研究', '国际交流', '国际交流', '国际会计', '动态', '党建', '工作', '\t', '\t', '您', '当前', '所', '位置', '委员会', '企业', '企业', '\t', '用', '案例', '实施', '问答', '实务', '问答', '趋同', '等效', '国际会计', '动态', '会计', '国际交流', '合作', '计', '人才培养', '企业', '企业', '解释', '其他', '规定', '企业', '第', '25', '号', '一', '一', '20202018', '会计', '司', '企业', '第', '25', '号', '一', '一', '财会', '[', '2020', ']', '20', '号', '第一章', '总则', '量', '相关', '信息', '列报', '根据', '企业', '一', '一', '基准', '则', '制定', '准则', '第二条', '指', '企', '人', '约定', '特定', '事项', '对', '保单', '持有人', '产生', '利', '影响', '给予', '其', '赔偿', '并', '因此', '有人', '重大', '风险', '事项', '指所', '承保', '产生', '风险', '确定', '未来', '事项', '风险', '指从', '保', '发', '人除', '金融风险', '之外', '风险', '第三条', '准则', '适用', '于', '下列', '一', '企业', '签发', '含', '三', '企业', '转让', '非', '同一', '控制', '下', '企业', '合并', '中', '取得', '上述', '签发', '企业', '所', '参', '分红', '特征', '投资', '适用', '准则', '指分入', '人', '签发', '人', '分出', '人', '约定', '对', '分', '引起', '赔付', '等', '进行', '补偿', '具有', '相机', '参', '分红', '特征', '投资', '指', '赋予', '特定', '保证', '金额', '附加', '金额', '金融工具', '附加', '金额', '由', '企业', '签发', '人', '基于', '特定', '项', '且', '预计', '构成', '利益', '重要', '部分', '第四条', '下列', '各项', '适用', '其他', '相关', '一由', '企', '一', '无形资产', '企业', '第', '14', '号', '一', '一', '收入', '企业', '第', '21', '号', '一', '一', '租赁', '项目', '未来', '使用', '情况', '等', '形成', '权利义务', '分别', '适用', '企业', '第', '6', '号', '一', '一',

```
i = 0
wordcount = {}
for word in Textsplit:
    i = i + 1
    wordcount[word]=wordcount.get(word,0)+1
    print(i, len(word)," ", word," ", wordcount[word])
print(i, wordcount)
```

4838	2	报告	1
4839	2	分部	1
4840	1	等	7
4841	1	对	36
4842	2	信息	6
4843	2	披露	2
4844	2	进行	33
4845	2	恰当	2
4846	2	汇总	3
4847	2	第八	2
4848	3	十四条	1
4849	2	企业	192
4850	2	资产	36
4851	3	资产负债表	14
4852	1	中	35
4853	2	分别	6
4854	2	列示	2

```

items = list(wordcount.items())
items.sort(key=lambda x: x[1], reverse=True)
text1 = ""
for i in range(len(items)):
    #print(items[i])
    word, count = items[i]
    if len(word)>1 and len(word)<5:
        if word == "期初": word = "期初余额"
        if word == "资产": word = "资产负债表"

```

```

print(word, count)
text1 = text1 + word + " "

```

```

WC = WordCloud(background_color="white",font_path = 'C:\Windows\Fonts\SimYou.ttf').generate(text1)
rcParams["figure.figsize"] = (12,4)
plt.imshow(WC)
plt.axis('off')
plt.show()

```



```
d = collections.Counter(wordcount)
text = " "
for word, count in d.most_common(25):
    print(word, ":", count)
    text = text + word + " "
```

Text=str(Textsplit)

合同	581
企业	271
现金流量	132
确认	123
服务	123
相关	103
变动	99
价值	97
金额	86
当期	81
分出	78
准则	77
资产负债表	77
责任	73
调整	73
亏损	71
计入	65
发生	62

```
d = collections.Counter(wordcount)
Record = ""
filename="D:/lzj/Word1.csv"
f = open(filename, "w")
for word, count in d.most_common(50):
    if len(word)>1 and len(word)<5:
        Record = word + "," + str(count)
        Record = Record.replace(" ", "")
        #print(Record)
        f.write(Record + "\n")
f.close()
```

1	合同	581	
2	企业	271	
3	现金流量	132	
4	确认	123	
5	服务	123	
6	相关	103	
7	变动	99	
8	价值	97	
9	金额	86	
10	当期	81	
11	分出	78	
12	准则	77	
13	资产	77	
14	责任	73	
15	调整	73	
16	亏损	71	
17	计入	65	
18	发生	62	
19	边际	62	
20	负债	61	

---

```
##2006 2009 旧保险合同准则

%matplotlib inline
%config InlineBackend.close_figures = False
import numpy as np
import matplotlib.pyplot as plt

from matplotlib import rcParams
from PIL import Image

import urllib
import bs4
from bs4 import BeautifulSoup

from wordcloud import WordCloud,STOPWORDS
import collections
import jieba

text=my_url = "file:D:/lzj2/jiu.txt"
wp = urllib.request.urlopen(my_url)
soup = BeautifulSoup(wp,'lxml')
text = soup.get_text().strip()
print(text)

text = text.replace(",","")
text = text.replace(";","")
text = text.replace("!","")
text = text.replace("?","")
text = text.replace("<","")
text = text.replace(">","")
text = text.replace("(","")
text = text.replace(")","")
text = text.replace("。","")
text = text.replace(":","")
text = text.replace(":","")
text = text.replace(" ","")
text = text.replace("的","")
text = text.replace("了","")
text = text.replace("在","")
text = text.replace("是","")
text = text.replace("也","")
text = text.replace("\n","")
text = text.replace("、","")
text = text.replace("\xa0","")
text = text.replace("—","")
text = text.replace(">","")
text = text.replace("\u3000","")
text = text.replace("\r","")
text = text.replace("《","")
text = text.replace("》","")
text = text.replace(",","")
text = text.replace("有","")
text = text.replace("我们","")
text = text.replace("和","")
text = text.replace("保险","")
text = text.replace("应当","")
text = text.replace("确定","")
text = text.replace("相关","")
text = text.replace("原","")
```

---

```

text = text.replace("为", "")
text = text.replace("与", "")
text = text.replace("人", "")
text = text.replace("—", "")
text = text.replace("未", "")
text = text.replace("再", "")
text = text.replace(".", "")
text = text.replace("时", "")
text = text.replace("已", "")
Textsplit = jieba.lcut(text)
print(Textsplit)

```

企业会计准则第 25 号——原保险合同

第一章 总则

第一条 为了规范保险人签发的原保险合同的确认、计量和相关信息的列报, 根据《企业会计准则——基本准则》, 制定本准则。

第二条 保险合同, 是指保险人与投保人约定保险权利义务关系, 并承担源于被保险人保险风险的协议。保险合同分为原保险合同, 是指保险人向投保人收取保费, 对约定的可能发生的事故因其发生所造成的财产损失承担赔偿责任, 或者被保险人达到约定的年龄、期限时承担给付保险金责任的保险合同。

第三条 下列各项适用其他相关会计准则:

(一) 保险人签发的原保险合同产生的损余物资等资产的减值, 适用《企业会计准则第 1 号——存货》。

(二) 保险人向投保人签发的承担保险风险以外的其他风险的合同, 适用《企业会计准则第 22 号——金融工具确认和计量——金融工具列报》。

(三) 保险人签发、持有的再保险合同, 适用《企业会计准则第 26 号——再保险合同》。

第二章 原保险合同的确定

'将分', '寿险', '非', '寿险', '延长', '期内', '承担', '赔付', '金', '责  
 '非', '寿险', '延长', '期指', '投保', '自上', '一期', '保费', '到期日',  
 三章, '收入', '第七条', '保费', '收入', '同', '满足', '下列', '条件', '二  
 '二', '经济', '利益', '很', '可能', '流入', '三', '收入', '能够', '可靠  
 '收入', '金额', '一', '对于', '非', '寿险', '根据', '约定', '保费', '总  
 '应', '收取', '保费', '一次性', '收取', '保费', '根据', '一次性', '应',  
 '应', '退还', '投保', '金额', '作退', '保费', '计入', '当期', '损益', '保  
 金', '决', '赔款', '准备金', '寿险', '责任', '准备金', '长期', '健康险',  
 '寿险', '责任', '提取', '准备金', '决', '赔款', '准备金', '指非', '寿  
 险', '责任', '准备金', '指尚', '终止', '寿', '责任', '提取', '准备金', '十  
 '责任', '提取', '准备金', '第十一条', '确认', '非', '寿险', '保费', '收  
 金', '作当期', '保费', '收入', '调整', '并', '确认', '到期', '责任', '保

```

i = 0
wordcount = {}
for word in Textsplit:
    i = i + 1
    wordcount[word]=wordcount.get(word,0)+1
    print(i, len(word), " ", word, " ", wordcount[word])
print(i, wordcount)

```

1	2	企业	1
2	4	会计准则	1
3	1	第	1
4	2	25	1
5	1	号	1
6	3	第一章	1
7	2	总则	1
8	3	第一条	1
9	2	规范	1
10	2	签发	1
11	2	确认	1
12	2	计量	1
13	2	信息	1
14	2	列报	1
15	2	根据	1
16	2	企业	2
17	4	会计准则	2
18	4	基本准则	1

```

items = list(wordcount.items())
items.sort(key=lambda x: x[1], reverse=True)
text1 = ""
for i in range(len(items)):
    #print(items[i])
    word, count = items[i]
    if len(word)>1 and len(word)<5:
        if word == "期初": word = "期初余额"
        if word == "资产": word = "资产负债表"

    print(word, count)
    text1 = text1 + word + " "

```

```

WC = WordCloud(background_color="white",font_path = 'C:\Windows\Fonts\SimYou.ttf').generate(text1)
rcParams["figure.figsize"] = (12,4)
plt.imshow(WC)
plt.axis('off')
plt.show()

```



```

d = collections.Counter(wordcount)
text = " "
for word, count in d.most_common(25):
    print(word, ":", count)
    text = text + word + " "

```

```
Text=str(Textsplit)
```

```

准备金 : 143
责任 : 75
分保 : 44
当期 : 41
风险 : 38
计量 : 36
寿险 : 35
赔款 : 34
发生 : 32
金额 : 31
保费 : 30
按照 : 30
赔付 : 29
费用 : 28
确认 : 27
分出 : 27
计入 : 25
损益 : 25
计算 : 24
根据 : 23
到期 : 23
重大 : 23
. : 23
约定 : 21
假设 : 20

```

```
d = collections.Counter(wordcount)
Record = ""
filename="D:/lzj2/Word2.csv"
f = open(filename, "w")
for word, count in d.most_common(50):
    if len(word)>1 and len(word)<5:
        Record = word + "," + str(count)
        Record = Record.replace(" ", "")
        #print(Record)
        f.write(Record + "\n")
f.close()
```

1	合同	148
2	准备金	143
3	责任	75
4	分保	44
5	当期	41
6	风险	38
7	计量	36
8	寿险	35
9	赔款	34
10	发生	32
11	金额	31
12	保费	30
13	按照	30
14	赔付	29
15	费用	28
16	确认	27
17	分出	27
18	计入	25
19	损益	25