



蚂蚁学习爱智慧



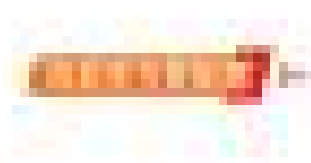
A N T S T U D Y

# 用大数据写好 财务报告

刘国红◎ 编著



重庆大学 电子音像出版社



人力资源社会保障部 财政部

# 用人单位写好 财务报告

人力资源和社会保障部  
财政部 编

## 内容提要

公文种类繁多，本书选取工作报告作为样本研究，分为理论模型与实操方法。理论模型部分，利用大数据技术，对国务院政府工作报告进行文本挖掘，建立总结与部署、开头与结尾、感谢与问题、体会与要求四个模块八大要素写作模型，分类提取高频词汇，总结提炼工作报告写作步骤及要点。实操方法部分，运用 Python、PHP、Mysql、Sqlite 等技术，构建个人知识库、公文语料库网站，提供全部源码，实现相关功能。

书名：用大数据写好财务报告

编著：刘国红

重庆大学电子音像出版社

责任编辑：陈 艳

地址：重庆市沙坪坝区大学城西路 21 号

ISBN 978-7-89446-766-9

出版时间：2023 年 8 月

邮编：400030

电话：023-88617079

字数：462 千字

定价：28.00 元

版权所有，侵权必究

年年岁岁花相似，岁岁年年人不同。如何写出一份让人眼前一亮的工作总结呢？人工智能与大语言模型已深刻影响到写作领域。用大数据写工作报告是一条可行的路子。结合近年来起草财务工作报告的经验，探讨几个普遍关注的话题。

## 一、财务总结为谁而写

不同岗位的财务总结是写给谁看的？这其实是在探讨财务总结的客体问题。季羨林老先生写过篇小品文《自己的花是给别人看的》，收入到小学课本。鲜花为欣赏而开，也给生命带来无限希望，但总结是为了更好地前行。自然界彰显了总结的魅力，竹子向上一段就小结一次，树木历经一年就向外拓展一圈，这都是总结的自然语言。孔子在《论语·述而》中提出，三人行，必有我师。体会财务总结至少要给三个人看，上级、自己和后来者。纵向看，财务总结给管理者看，了解问题、推动进展、评价绩效；横向看，财务总结给自己看，梳理心得、看清得失、提升自我；回头看，财务总结给后来者看，镌刻过往、以史为鉴、与时俱进。柳宗元在《南涧中题》中讲，谁为后来者，当与此心期。说的就是这个意思。起草文字材料，经常有似曾相识的感觉，就像歌词里唱的，我们好像在哪见过，你记得吗？喔，原来你也在这里。

## 二、横看成岭侧成峰

哪个岗位的财务总结最值得关注？这是在讨论财务总结的本体问题。每个人都是独一无二的存在，在不同的岗位能领略到独特的风景。从会计重要性原则看，有些岗位的财务总结的确需要更多关注，比如，承接的重点任务、考核的关键指标、攻坚的改革事项，这些工作具有标志性、关联性、引领性，是牛鼻子、牵

一发动全身，是撬字诀、一子落满盘活。

作为接收者，财务总结怎么看？最容易忽视的关键点是什么？这是在关注财务总结的主体问题。苏轼在《题西林壁》中描写庐山，横看成岭侧成峰，远近高低各不同。这告诉我们，看世界的角度不同，看到的世界也不尽相同。大家都知道，毛主席评价《水浒传》好就好在投降，其实后面还有两句话，作反面教材，使人民都知道投降派。财务工作总结是从理论高度认识过去，从实践角度指导未来，概括起来有六看一回顾。一看发现了什么问题，二看形成了什么思路，三看提出了什么措施，四看提炼了多少亮点，五看创造了多少业绩，六看制定了什么新目标，一回顾就是当前的结果与起初的计划一致吗？

俗话说，不看鱼面看水面。财务工作总结也有暗流涌动，也有波澜不惊。从指标上看，容易关注利润，而忽视现金流，很多账面盈利多的企业寿命却很短。从章法上看，转折句前后的内容都是关键点，摆成绩多看定语，讲困难多看补语。

业财融合背景下，财务总结除了财务部门的事情之外，是否应该覆盖业务，比例大概是多少？这是在探寻财务总结的共同体问题。影子离不开光，光离了影子却毫发无伤。会计四大假设就有持续经营，没有业务，财务也没有独立存在的意义。财务伴随着业务，数字背后都是经济活动，财务尽可能用专门语言描述经济运行轨迹，如有可能也会揭示其内在运行规律。财务总结涉及多少业务成分，取决于业财融合的程度，也反映了财务对业务的支撑力度。

### 三、精彩总结眷顾有准备的人

抛开业绩、工作量，优秀的财务工作总结是什么样的？具备哪些元素？这是在探究财务总结的载体问题。托尔斯泰在《安娜·卡列尼娜》开篇就指出，幸福的家庭都是相似的，不幸的家庭各有各的不幸。大多财务总结涉及商密，一般不容易看到全文，大家有兴趣可以看中国会计年鉴，或者各上市公司年报，里面都涉及到财务总结。好的财务总结让人产生共鸣，加上各自感悟越读越长，反之，味如嚼蜡越读越短。

从历史长河看，诸葛亮写的《出师表》堪称千古一流的应用文，先帝创业未半而中道崩殒，今天下三分，益州疲弊，此诚危急存亡之秋也。开篇就是精准到位的形势分析。亲贤臣，远小人，此先汉所以兴隆也；亲小人，远贤臣，此后汉所以倾颓也。这是入木三分的根源剖析，为我们提供了很好的分析范本。

从谋篇布局看，骆宾王写的《咏鹅》别出心裁，天衣无缝。“鹅，鹅，鹅，曲项向天歌。白毛浮绿水，红掌拨清波。”一句只有三个字，这种体例在五言绝句中罕见，但动静相生，活泼生动，十分传神。与此有异曲同工之妙的还有汉高祖刘邦写的《大风歌》，大风起兮云飞扬，威加海内兮归故乡，安得猛士兮守四方！一首诗只有三句，

这在中华文化史上也不多见，总体上大气磅礴、别具一格。

从行文逻辑看，写文章一般会讲三点以上，就像盖房子，既要总体框架，也要有起承转合，逻辑好比楼梯，从底层到顶层要通畅。经常讲文章逻辑性不强，其实相当于从一楼到三楼，二楼没有楼梯，论据得不出相应结论，文章看起来就显得牵强附会。

从遣词造句看，文章让人过目难忘，还有一个因素就是文字灵动不刻板。民间有鲜活的语言。比如，花鼓戏《站花墙》（花墙会）剧本，台词画面感强、神韵十足。有时我们感觉工作无话可说，《站花墙》用6个排比句写出了万苦千辛，干了什么就说什么，很值得我们学习。“我为你千里奔波冒风尘，我为你死里余生血染巾，我为你挨过王府无情棍，我为你含悲忍辱入空门，我为你墙外脚印擦脚印，我为你手拿木鱼敲碎心。”

讲了这么多，好的财务总结到底要具备什么要素呢？吕坤在《呻吟语》中讲，文章有八要：简、切、明、尽、正、大、温、雅。不简则失之繁冗，不切则失之浮泛，不明则失之含糊，不尽则失之疏遗，不正则理不足以服人，不大则失冠冕之体，不温则暴厉刻削，不雅则鄙陋浅俗。庙堂文要有天覆地载，山林文要有仙风道骨，征伐文要有吞象食牛，奏对文要有忠肝义胆。这对撰写财务总结有很强的现实参考意义。

年度财务总结应该从什么时候开始着手？起草财务工作报告有高效方法吗？好的文章需要时间酝酿，就像优秀的人值得等待。从实际操作看，每年总结财务工作从当年10月份就开始准备了，一对前三季度工作有了基本掌握，二对第四季度工作也有大体预测，三要召开管理务虚会研讨下年工作，基本具备了总结全年财务工作的条件。

宋代理学家周敦颐提出文以载道，理解思想是文章的灵魂，思想性不强，经不住岁月的洗礼。大家都知道提高站位，这也是增强思想性的有效途径。

常言道，好的开始，是成功的一半。有个红烧肉做法的例子很经典，有人会就事论事，详尽写下每个烹饪步骤；也有人会追根溯源，从选好食材着手；还有人会透过现象看本质，从如何养猪开始写，这从侧面反映了怎么写文章的第一句话。我们经常说，人无常师，水无常形，兵无常势，文无定法，但也有约定成俗的架构。总结的总字上面有两点，一个是重点、一个是亮点，写重点提纲挈领，就像撒网，由面到点，干货满满；讲亮点纲举目张，类似撑伞，由线到面，辐射一片，这样财务总结就会出力更出彩。财务总结通常分为三部分，一是工作回顾，二是形势分析，三是下步安排。工作总结采取回溯法，总结年度看半年，总结半年看季度，总结季度看月度，总结月度看周记，只有平时工作日志详实，根据岗位职责和年度计划梳理出大事件相对容易，工作回顾要逻辑清晰、数据支撑、突出成绩。形势分析多关注全球大势，多浏览专业

期刊，对未来走势看得准不准，直接影响后续部署。下步安排大多是结合当前形势和任务提出的，针对性会很强。

好的财务总结到底是怎么练成的呢？熟读唐诗三百首，不会吟诗也会吟。这说明多见多识广下笔如有神。扬州瘦西湖饯春堂有副楹联：莺啼燕语芳菲节，蝶影蜂声烂漫时。这里摘取了唐五代诗人毛熙震、李建勋诗词，把春天的听觉、嗅觉、视觉尽收眼底。写文章也是一样，有充足的样本，建立个人知识库，形成结构化的字词句篇大数据，终究会有山花烂漫时。信息化技术迅猛发展，为财务工作总结提供了数字化思维和工业化工具。把日常获取素材的网站分门别类整理，采取前中后台管理，后台用 Python 定期批量爬取资讯，中台用 sqlite 数据库存储资源，前台用 delphi 展示检索数据，尽管技术不是很前沿，但为个人写作提供了很大便利。

如果天下无新事，就像《圣经》里讲的，已有的事、后必再有，已行的事、后必再行，日光之下并无新事。写好财务总结，字词句篇数据库就派上大用场了，把公司成立以来的财务工作报告数字化，存储到数据库，按句子拆分，分为篇名、标题、开头、结尾、过渡句等类型，可以按音序检索，无论你想写什么，总能从历史中寻找印记，找到很多所见略同的样本做参考，是谁开始先出招没什么大不了，可借鉴、可改造、可创新，的确做到手中有粮、心里不慌。

如果天下有新事，就像鲁迅在《少年闰土》中讲的，我素不知道天下有这许多新鲜事。把唐诗宋词元曲、《人民日报》《求是》、各级政府工作报告、党代会报告等素材建立数据库，也是按句子检索，有两个发现，一是有很多句子重复率很高，二是样本数量超过百万级数据库运算有得天独厚的优势，跟简单运用 Word、Excel 不可同日而语，当你看过了数以万计的标题，每一个词都有数千种遣词造句，就会由衷感叹数据库让天下没有难写的财务总结。只要用心勤学苦练方法对路，每个财会人都能写好工作总结。

建立个人知识管理体系是一个长期积累的过程，利用数字化工具建立数字化知识库可达到事半功倍。在实践中，探索形成蚂蚁学习法。AntStudy 简称蚂蚁学习法，由 ANT 学习模型与 Stud ( x , y ) 成长函数两部分构成。ANT 学习模型旨在提升工作能力，构建个人知识库，分为吸收 ( Absorb )、叙述 ( Narrate )、训练 ( Train ) 三个关键环节。Stud ( x , y ) 成长函数侧重调整工作心态，遵循大树成长规律，用数学语言提炼出成长可变函数，从萌芽到栋梁需调动向阳、时间、向上、扎根、渴望五个变量。一是向阳 ( Sun )。心向阳光、向光生长，才能争取更多光明。二是时间 ( Time )。岁月刻画年轮，树干一圈圈往外长，不断得到积累经验，人的成长也要经历时间磨砺。三是向上 ( Up )。树先长主干再长细枝，越向上长，空间越大。人的心态也要阳光豁达，

受得了委屈，沉得住气。四是扎根（Down）。大树扎根大地，吸取营养，充实自我，牢固才能经得起岁月和时间的考验。人的成长也要有擅长的领域，片技即足自立天下。五是渴望（Yearn）。这是一个 X 与 Y 可选项。X 代表 Stop 以静制动，Y 代表 Move 以动制静。树挪死、人挪活，但大树往往选择经风霜、历风雨、屹立坚守。

刘国红

2023 年 1 月 21 日

于北京朝阳门

绪 论 / 01

第一篇 吸收 (Absorb) / 09

第一章 总结与部署词汇簇及语料库 / 11

第一节 国务院政府工作报告总结与部署文本挖掘 / 13

第二节 各省市政府工作报告总结与部署文本挖掘 / 36

第二章 开头与结尾词汇簇及语料库 / 121

第一节 国务院政府工作报告开头与结尾文本挖掘 / 122

第二节 各省市政府工作报告开头与结尾文本挖掘 / 135

第三章 感谢与问题词汇簇及语料库 / 170

第一节 国务院政府工作报告感谢与问题文本挖掘 / 171

第二节 各省市政府工作报告感谢与问题文本挖掘 / 190

第四章 体会与要求词汇簇及语料库 / 236

第一节 国务院政府工作报告体会与要求文本挖掘 / 238

第二节 各省市政府工作报告体会与要求文本挖掘 / 247

第二篇 叙述 (Narrate) / 295

第一章 词汇簇首字检索 / 297

第二章 词汇簇尾字检索 / 420

第三篇 训练 (Train) / 545

第一章 财务工作报告总结与部署 / 547

第二章 财务工作报告开头与结尾 / 551

第三章 财务工作报告感谢与问题 / 553

第四章 财务工作报告体会与要求 / 555

参考文献 / 558

## 绪 论

### 一、研究背景

语言是一个复杂的层次系统。语音、词汇、语法被称为传统语言学三要素。词语是语言的最小单位，是文章的基本要素。句子是表达完整思想的语言单位，句子由词语构成。词语与句子相互依存，排列组合形成字词句篇，表达清晰的意思。

现代文学家鲁迅写的《从百草园到三味书屋》，开篇很是经典。“不必说碧绿的菜畦，光滑的石井栏，高大的皂荚树，紫红的桑椹；也不必说鸣蝉在树叶里长吟，肥胖的黄蜂伏在菜花上，轻捷的叫天子（云雀）忽然从草间直窜向云霄里去了。单是周围的短短的泥墙根一带，就有无限趣味。油蛉在这里低唱，蟋蟀们在这里弹琴。翻开断砖来，有时会遇见蜈蚣；还有斑蝥，倘若用手指按住它的脊梁，便会啪的一声，从后窍喷出一阵烟雾。”这段文字不足 200 字，接连运用形容词与动词，极为生动地描写了童真童趣。

宋代词人辛弃疾写的《西江月·夜行黄沙道中》，用词也很是考究。“明月别枝惊鹊，清风半夜鸣蝉。稻花香里说丰年，听取蛙声一片。七八个星天外，两三点雨山前。旧时茅店社林边，路转溪头忽见。”这首词中的“说丰年”与“听蛙声”画面感十足。有人模仿写出，站在一望无垠的稻田，看到扑面而来的滚滚麦浪。用词不妥，就会漏洞频出，稻田里看麦浪显然是鸡同鸭讲。

公文是一种实用文体。公文写作研究涵盖写作理念、写作内容、写作格式、写作语言、写作主体、写作教学、处理流程等领域，体现了制度化、规范化、科学化趋势，量化分析是自然语言分析的发展方向。公文语体是具有明确的

交际对象、实施效能和固定程式的书面语言系统。以工作报告最为典型，是代表集体意志的大文章，写作特点、逻辑结构、格式要求具有特殊性和规范性，以叙述为主，多用陈述句和祈使句，行文庄重，用词讲究。陈述句是表述情况、说明问题、评议现象的主要方式。祈使句具有特定的实施效能，根据所用动词的不同，可以分为命令型、能愿型、祈望型和禁止型。命令型常用“命令、令、号召”等动词。能愿型常用“应该、要、必须”等能愿动词。祈愿型常用“请、希望、建议”等动词。禁止型常用“禁止、不准、不得”等动词。动词是工作报告活的灵魂，具有强大的组篇功能，选用适当动词造句，围绕观点集句成段，紧扣逻辑框架组段成篇，对工作报告写作具有借鉴意义和参考价值。

## 二、样本选择

各级人民代表大会和中国人民政治协商会议统称两会。两会是中国政治生活的大事，是关注中国最新发展政策的重要窗口，其中重头戏就是讨论五篇报告，即中国人民政治协商会议全国委员会常务委员会工作报告、全国人民代表大会常务委员会工作报告、一府两院的工作报告。这五份工作报告是“经国之大业”的重要文稿，是总结过去的成绩单，也是规划未来的行动纲领，与国计民生高度关联，具有极强的指导性。本书以政府工作报告为研究样本，基于政府工作报告具有三个鲜明特征。一是框架固定。正文一般包括工作总结、工作任务和自身建设等内容。开头与结尾部分有基本格式。二是内容全面。涵盖政治、经济、文化、社会、生态、文明建设等诸多领域，每一项工作均有关注群体。三是高度概括。每一句话都有深刻的内涵，每一项政策措施背后都有丰富的具体内容，重点任务会逐项分解落实。

## 三、研究方法

语料库，即语言资料库，是为特定应用目的采集的具有结构化、代表性、可检索的规模化语言文字集合。语料库是承载语言知识的基础素材，也是语言学的重要资源，是自然语言处理的基本要件。本书基于语料库开展工作报告写作研究。

一是建立报告语料库。以国务院工作报告为基本文本，增加各省市部分政府工作报告纳入语料库。收录国务院政府工作报告共约 119 万字，省

市政府工作报告约 311 万字，本书建立的政府工作报告规模约为 430 万字。MySQL 是一种关系型数据库管理系统，多用于各种网站和应用程序。SQLite 是一款轻量级关系数据库系统，非常适合单机版应用程序。本书语料库应用 MYSQL、SQLite 数据库，便于搭建 MYSQL+PHP 工作报告网站，开发本地语料库管理程序。

二是构建模块语料群。大样本分析历届政府工作报告框架结构，将工作报告分解为总结与部署、开头与结尾、感谢与问题、体会与要求四个模块八个要素，用结构化思维开展专题量化研究，分项建立四个模块语料。受篇幅限制，本书未列举所有语料信息，相关数据提供电子版。

三是形成专项词汇簇。以语料库为底层依据，进行词语切分，分级统计国务院与省市市政府工作报告高频词与常用词，去重剔除，建立工作报告八个要素的专项词汇簇。本书提炼出 3535 个常用词语，后续可深化扩充。

四是贯通检索关联谱。建立词汇簇及语料库的映射，利用数据库可检索每个词语的语料出处，相当于词典的例句功能。按专项词表首字和尾字排序，便于查询组织排比句等特定句式，增强澎湃气势。

#### 四、基本规则

数据库是经过结构化加工存储的相关数据集合，要按数据结构进行组织、存储和管理。模型、模块、词汇簇、语料库是本书的重要研究要件。模块具有群集特性，在写作中具有聚拢作用。本书将工作报告分解为总结与部署、开头与结尾、感谢与问题、体会与要求四个模块八个要素进行专项解析，按数据库检索规则予以编码。

表 1-1 工作报告写作模块

| 项目   | 模块 I |    | 模块 II |    | 模块 III |    | 模块 IV |    |
|------|------|----|-------|----|--------|----|-------|----|
|      | 序号   | 要素 | 序号    | 要素 | 序号     | 要素 | 序号    | 要素 |
| 工作报告 | 1    | 总结 | 3     | 开头 | 5      | 感谢 | 7     | 体会 |
|      | 2    | 部署 | 4     | 结尾 | 6      | 问题 | 8     | 要求 |

表 1-2 工作报告词汇簇编码规则

| 序号 | 大类 | 类别 | 词语 | 频次 | 高频词 | 语料号 | 种类 | 年份 | 检索号 | 首字 | 尾字 |
|----|----|----|----|----|-----|-----|----|----|-----|----|----|
|    |    |    |    |    |     |     |    |    |     |    |    |

序号,按语料量级编号,可按4位数排序或按大类+4位数排序。

大类,词语所属大类1总结、2部署、3开头、4结尾、5感谢、6问题、7体会、8要求,存在一个词同属多个大类的情景。

类别,用2位数表示。十位数1代表国务院,十位数2代表省市。个位数使用规则与大类相同。11国务院总结、21省市总结、12国务院部署、22省市部署、31国务院开头、32省市开头、41国务院结尾、42省市结尾、51国务院感谢、52省市感谢、61国务院问题、62省市问题、71国务院体会、72省市体会、81国务院要求、82省市要求。

词语,根据工作报告语料分词获取。

频次,统计词语在语料库出现次数。

高频词,用2位数表示,十位数使用规则参考大类编码。10总结高频词、20部署高频词、30开头高频词、40结尾高频词、50感谢高频词、60问题高频词、70体会高频词、80要求高频词。

语料号,由3个字母+4位数构成。比如,GWY0001代表国务院第1号工作报告语料,BJS0003代表北京市第3号工作报告语料。字母编码规则GWY国务院、BJS北京、TJS天津、HBS河北、SXS山西、NMG内蒙古、LNS辽宁、JLS吉林、HLJ黑龙江、SHS上海、JSS江苏、ZJS浙江、AHS安徽、FJS福建、JXS江西、SDS山东、HNS河南、HBS湖北、HNS湖南、GDS广东、GXS广西、HNS海南、CQS重庆、SCS四川、GZS贵州、YNS云南、XZS西藏、SXS陕西、GSS甘肃、QHS青海、NXS宁夏、XJS新疆。

种类,段落语料所属大类,规则与大类相同。1总结、2部署、3开头、4结尾、5感谢、6问题、7体会、8要求。

年份,工作报告年份用4位数表示。

检索号,指语料检索号,由种类+年份+语料号共12位数构成。比如,12023GWY0001代表国务院工作报告2023年总结类第1号语料。

首字,提取词语左起第一个汉字。

尾字,提取词语右起第一个汉字。

表 1-3 工作报告语料库编码规则

| 序号 | 语料号 | 种类 | 年份 | 检索号 | 内容 |
|----|-----|----|----|-----|----|
|    |     |    |    |     |    |

语料库编码规则与词汇簇编码规则相同。

## 五、研究内容

本书以工作报告为研究场景,以历届国务院政府工作报告及近五年全国各省市政府工作报告文本为语料素材,搭建ANT学习模型,探讨工作报告写作规程,分为吸收、叙述、训练三个篇章阐述。

第一篇,吸收(Absorb)。根据专业方向,选择相关知识,进行系统消化,

吸收知识精髓，提炼核心观点，为搭建个人知识框架累积素材。第一章提炼总结与部署词汇簇，统计高频词及常用词。第二章提炼开头与结尾词汇簇，统计高频词及常用词。第三章提炼感谢与问题词汇簇，统计高频词及常用词。第四章提炼体会与要求词汇簇，统计高频词及常用词。

第二篇，叙述（Narrate）。知识能意会可言传，更有利于传播共享创新。叙述就是精确分类，对入库知识打标签，为深度检索奠定基础。本篇建立工作报告词汇簇及语料库的关联，第一章按首字检索，第二章按尾字检索，便于统一句式字，形成排比句。

第三篇，训练（Train）。实操是最好的检验，解决问题是构建个人知识库的根本目的，在训练中提升改进，在训练中调整完善。第一章运用总结与部署词汇簇，形成财务工作报告提纲。第二章运用开头与结尾词汇簇，生成财务工作报告开头与结尾部分文字内容。第三章运用感谢与问题词汇簇，构成财务工作报告感谢文本与列示问题。第四章运用体会与要求词汇簇，组成财务工作报告实践体会与总体要求。

需要特别强调的是，工作报告安排字、词、句、段具有全局意识，准确理解内容和意图，不能只注重单独的某个词、某个句子，要在总体布局下把握大体框架及写作思路，切忌只见树木，不见森林。





吸尽百川还见底  
收拾清晖上笔端

