



# 论文查重平台简介与原理及降重方法

北京林业大学信息学院



汇报人：瞿小龙



汇报时间：2025年4月

# 目录

## CONTENTS

01

查重平台简介

02

论文查重规则

03

论文降重方法

04

查重常见问题



## 01、查重平台简介



# 主流查重平台的不同定位

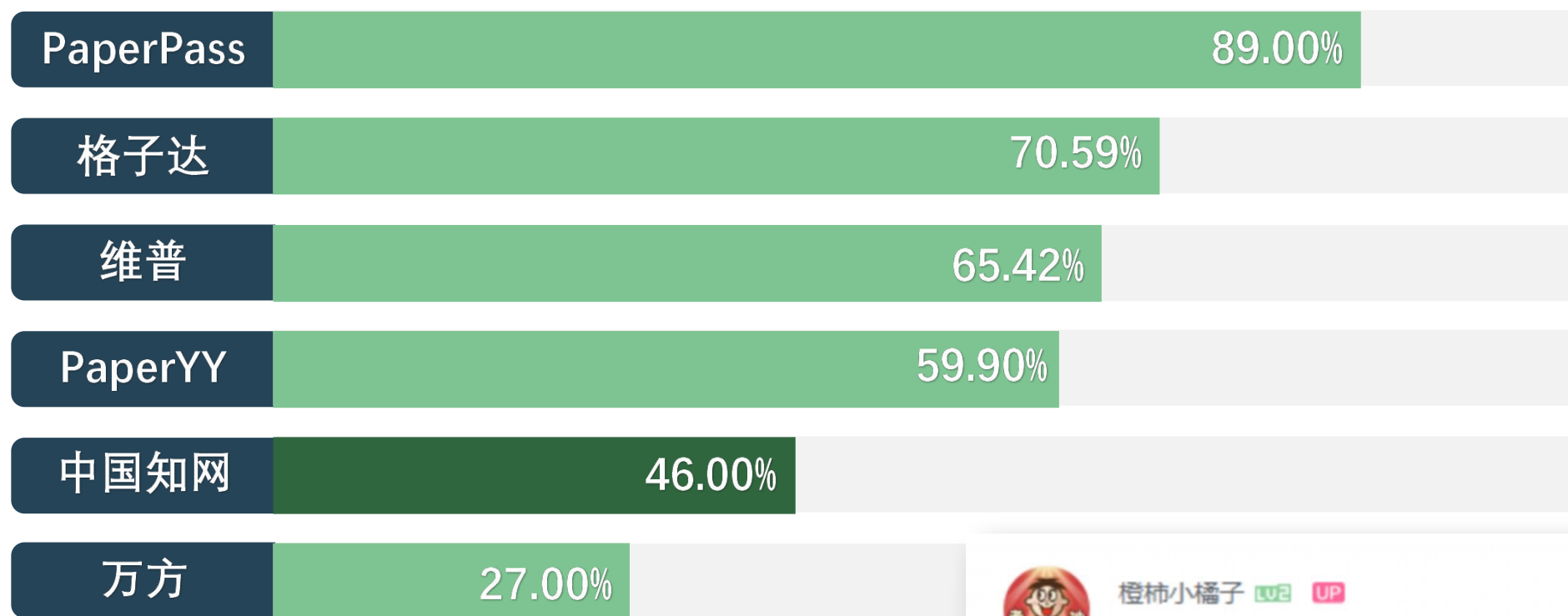
查重平台	高校认可 官方平台	提供个人 购买服务	提供免费 查重服务	价格 (元/千字符)
中国知网	✓	✓	—	按篇收费
维普	✓	✓	—	3
万方	✓	✓	—	3
格子达	✓	✓	—	3
Paper系*	—	✓	✓	0.3~1.5
freecheck	—	✓	✓	0.3~1.5
笔杆	—	✓	✓	1.5

注：Paper系查重平台有PaperPass、PaperYY、PaperAsk等



## 查重结果不一，知网分数偏低

### 21年（天临三年）翟天临硕士论文各平台查重结果对比



查重率不同的原因

- 数据库不一样
- 查重算法不同



橙柿小橘子 LV2 UP

1. 查重软件:

paperyy至尊版3次, 13.4%, 11%, 7.2%, 免费版2次, 12%, 6.2%;

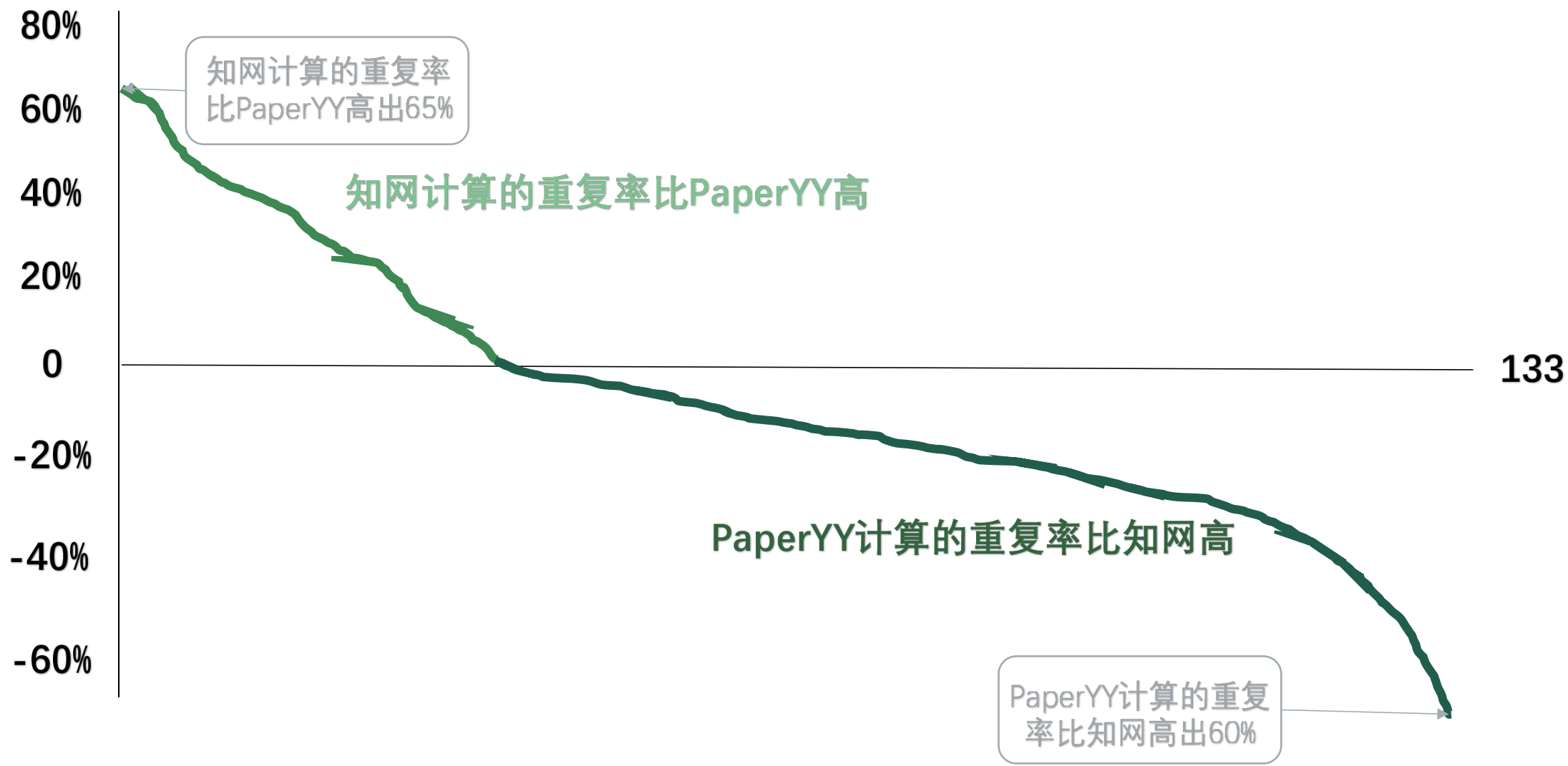
paperfree<sup>a</sup>一次23.82%;

学信网万方一次2.87%



## 查重结果不一，知网分数偏低

### 133位网友在知网和PaperYY的查重结果对比分析





## 02、论文查重规则



# 各平台区别：论文对比库

## 知网数据库

中国学术期刊库  
中国博士学位论文全文数据库  
中国优秀硕士学位论文全文数据库  
中国重要会议论文全文数据库  
中国重要报纸全文数据库  
中国专利全文数据库  
图书资源优先出版文献库  
大学生论文联合比对库  
互联网资源(包含贴吧等论坛资源)  
英文数据库港澳台学术文献库  
互联网文档资源  
源代码库  
原创作品库  
个人比对库

## 维普数据库

中外期刊数据库  
学术会议论文库  
硕博学位论文库  
特色自建论文库  
互联网数据资源  
维普系统检测文献库  
中文科技期刊论文全文数据库  
博士/硕士学位论文全文数据库  
外文特色文献数据全库  
中文主要报纸全文数据库  
中国主要会议论文特色数据库  
中国专利特色数据库  
维普优先出版论文全文数据库  
互联网数据库资源/互联网文档资源  
港澳台文献资源  
图书资源  
古籍文献资源  
IPUB原创作品  
年鉴资源

## 万方数据库

中国学术期刊数据库  
优先出版论文数据库  
国内外重要学术会议论文数据库  
中国博士学位论文全文数据库  
中国优秀硕士学位论文全文数据库  
中国优秀报纸全文数据库  
互联网学术资源数据库  
学术网络文献数据库  
中国专利文献全文数据库  
特色英文文摘数据库  
中国标准全文数据库

## Paperpass数据库

(免费版)  
中文学术期刊库  
中文学位(硕博)论文库  
中文会议论文库  
大学生联合比对库  
(旗舰版)  
中文学术期刊库  
中文学位(硕博)论文库  
中文会议论文库  
大学生联合比对库  
中文图书馆  
专利数据库  
报纸数据库  
外文数据库  
互联网资源  
论文纠错  
极速检测

差别还是非常大的





# 以知网为例：查重流程

## 1.上传论文

将要检测的论文上传到查重系统中。

## 2.文本处理

对上传的论文进行文本处理，例如去除格式、标点符号、换行符和空格等。

## 3.切割文本

将论文切割成若干个片段，一般大小为2-3个单词至数个句子不等。

## 4.建立索引

为论文中的每个片段建立索引，并将这些索引存储在数据库中。

## 5.查重规则

将上传的论文中的每个片段与数据库已有的索引按照平台的查重规则进行比较，并计算相似度。

## 6.生成报告

根据比较结果生成查重报告，比较出与其他论文相似的片段，并给出相似度分值和原文出处。



# 知网查重规则

## ①：分章节分段进行检测，查重系统阈值为5%。

论文要求是要**整篇上传检测**，论文上传到学术不端监测系统之后，该系统会根据你**论文的目录进行检测**，将论文分成章节来检测，这样既能看到每一单章节的复制比，同时能够保证目录不参与正文检测；否则，会**自动分段**按照**1万字符**左右检测，同时目录有可能当初正文检测，重复就会标红。





# 知网查重规则

## ①：分章节分段进行检测，查重系统阈值为5%。

论文要求是要**整篇上传检测**，论文上传到学术不端监测系统之后，该系统会根据你**论文的目录进行检测**，将论文分成章节来检测，这样既能看到每一单章节的复制比，同时能够保证目录不参与正文检测；否则，会**自动分段**按照**1万字符**左右检测，同时目录有可能当初正文检测，重复就会标红。

知网检测结果实例

原文内容		相似内容来源
1	此处有 58 字相似 小型化、易使用、低成本、零排放等优良[1]。电卡制冷被普遍认为是一种能够取代传统压缩制冷的制冷方式。本章节，将对铁电材料电卡制冷的基础理论及研究进展进行详细介绍。 <u>1.1 电卡效应的基础理论</u> <u>1.1.1 电卡效应的概念及制冷原理</u>	钛酸钡基陶瓷复合材料的电卡效应及储能特性研究 李明顶 - 《广东工业大学博士论文》- 2019-05-01 (是否引证：否) 1. ;Oxygen vacancies摘要 -9Abstract -7第一章 绪论 1 1.1 电卡效应的概念及研究进展 1 1.1.1 电卡效应的概念 1 1.1.2 制冷原理 1 1.1.3 研究进展 2 1.2 无铅铁电陶瓷的储能及研究进展 4 1.2.1 无铅铁电陶瓷
	4.9%	7210348 刘昊然 基于提示学习的数据结构课程知识图谱构建及问答研究.pdf 第1部分( 总10199字 )
	2.2%	7210348 刘昊然 基于提示学习的数据结构课程知识图谱构建及问答研究.pdf 第2部分( 总9049字 )
	2.8%	7210348 刘昊然 基于提示学习的数据结构课程知识图谱构建及问答研究.pdf 第3部分( 总9869字 )
	6%	7210348 刘昊然 基于提示学习的数据结构课程知识图谱构建及问答研究.pdf 第4部分( 总9338字 )
	0.3%	7210348 刘昊然 基于提示学习的数据结构课程知识图谱构建及问答研究.pdf 第5部分( 总9654字 )
	19.1%	7210348 刘昊然 基于提示学习的数据结构课程知识图谱构建及问答研究.pdf 第6部分( 总7087字 )





# 知网查重规则

## ①：分章节分段进行检测，查重系统阈值为5%。

论文要求是要整篇上传检测，论文上传到学术不端监测系统之后，该系统会根据你论文的目录来进行检测，将论文分成章节来检测，这样既能看到每一单章节的复制比，同时能够保证目录不参与正文检测；否则，会自动分段按照1万字符左右检测，同时目录有可能当初正文检测，重复就会标红。

致谢

1929-1936.

[18] 张萌. 课程知识图谱组织与搜索技术研究[D]. 武汉: 武汉大学, 2016.

[19] 杨文. 基于百度百科的知识图谱构建方法研究[D]. 湖北: 华中师范大学, 2019

致谢

68

[50] Ma R, Zhou X, Gui T, et al. Template-free prompt tuning for few-shot NER[C]. Proceedings of the

2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Washington, 2022.

[51] Han X, Zhao W, Ding N, et al. PTR: prompt tuning with rules for text classification[J]. AI Open, 2022, 3: 182-192.

[52] Chen X, Zhang N, Xie X, et al. KnowPrompt: knowledge-aware prompt-tuning with synergistic optimization for relation extraction[C]. The Web Conference 2022. France, 2022.

[53] Li X L, Liang P. Prefix-tuning: optimizing continuous prompts for generation prefix-tuning: opti-

mizing continuous prompts for Generation[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Thailand, 2021.

“致谢”参与查重，“参考文献”不参与查重。但由于“参考文献”上方的页眉不小心写成了“致谢”，导致“参考文献”也参与了查重，进而查重率直线上升。

在论文参考文献格式正确的前提下，知网查重系统不对参考文献查重，否则会被用来查重。



## ①：分章节分段进行检测，查重系统阈值为5%。

学术不端系统存在一个阈值5%，如果你论文中引用别人的片段不超过总体字数的5%的话，那么这些内容并不会被算作是重复的，这种情况常见于大段落中的小句或者小概念。举个例子：假如检测某个段落有10000字，那么引用单篇文献500字以下，是会被检测出来的。实际上这里也告诉同学们一个修改的方法，就是一个段落千万不要只选一篇文章来引用，尽可能多的选择多篇文献，一篇截取几句，这样是不会被检测出来的。

Neo4j作为高性能的NOSQL图形数据库，具有高性能、嵌入式和轻量级等优点。其主要优势如下：

- (1) 采用深度遍历进行查询，在大规模数据集中可以保持良好性能；
- (2) 支持包含Java、Python在内的多种程序语言访问，同时使用Cypher语句实现对图数据库的查询，操作实现容易；
- (3) 具有较好的灵活性，对环境的依赖较小。

### 2.4 本章小结

本章主要介绍了论文涉及的相关理论和技术，包含命名实体识别相关理论、集成学习相关理论和系统实现相关技术。根据研究内容，本文将采用BiLSTM-CRF和SoftLexicon-BiLSTM-CRF这两个NER模型，并结合集成学习相关理论以完成植物NER任务。同时，本文将选用后端Django框架、前端Vue框架和Neo4j图数据库以实现系统功能。

### 3 基于神经网络集成的植物命名实体识别模型

还有另外一个规则：连续13个字相似或者抄袭会被红字标注。



## ②：模糊计算，不以句子为最小检测单位。

系统的最小检测单位是句子吗？如果是，那么在每句话里改动一两个字就检测不出来了吗？

查重系统采用的是语义级别检测技术，没有“最小检测单位”的概念。系统在识别重复和引用内容时，会结合上下文的内容，对达到一定语义级别的内容进行鉴定，并不是单纯根据一两个词、字或者单独的句子进行判断。完整内容的重合情况，是由系统根据算法综合得出的，对文献内容的远样抄袭、该写、语句顺序调整等，都能自动检测和识别，且能快速定位和动态标注显示。

### 知网检测结果实例展示：

<p>此处有 36 字相似</p> <p>制冷剂绝热压缩。（2）第二阶段，此过程为施加恒定测试电场时的热量转移过程。在材料两端的电场保持恒定的条件下，通过方法</p> <p>热量被导出，此过程相当于传统蒸汽压缩制冷机中制冷剂的等温压缩。</p> <p>（3）</p>	<p>铁酸钡基陶瓷复合材料的电卡效应及储能特性研究 李明顶 - 《广东工业大学博士论文》 - 2019-05-01（是否引证：否）</p> <p>1. 类似。（2）过程二为恒定的外加电场下的热量转移过程：即保持过程一中施加的电场不变，使得铁电材料与散热装置接触，将热量转移出去。此过程与传统空气压缩制冷中制冷剂的等温压缩相类似。（3）过程三为磁电材料下的去极化过程，即在磁电材料下，撤除过程</p>
---	---





## ③：连续13个字符相同或者相似，则判为重复

查重系统采用“句群语义模糊算法”：关注句子中心思想和关键词，不会记入虚词（之乎者也）；同时，系统也注重“二次检测”：如果语句被判定为疑似重复，系统会加强寻找和该语句相关的内容，并再次进行比对。

其他平台：维普：如果一个句子中有连续10~20个字出现重复，则判定句子涉嫌抄袭。Paper系：一个句子中超过五个字及以上所有关键字全标黄。

## ④：声明、目录、参考文献不参与查重，但“致谢”作为正文检测。

声明、目录、参考文献、致谢等内容随正文一同上传即可，系统可以自动判断该部分内容，并进行相应处理。

声明、目录和参考文献会自动排除，不作为正文进行查重查重，并根据目录和参考文献，判断论文的分段及引用情况；致谢作为正文，参与查重检测。

## ⑤：公式、表格会被查重检测，图片不会被检测。

中国知网公告：学术不端的各种行为中，文字复制是最为普遍和严重的，目前本检测系统对文字复制的检测已经达到相当高的水平。同时，系统已经实现了对公示和表格的检测，并将有关的检测结果展示在检测报告中。而对于图片内容的检测，目前系统已经具备了图片检测的技术，正在进行集成的测试和优化。



## 03、论文降重方法





# 论文降重可行方法总结

低难度

## 中外语言互译法

先把中文翻译成外文，再把外文翻译成中文。  
这是一种很古老的方法，据说英语效果最差，  
白俄罗斯语（小语种）效果较佳。

## 表文/图文/图表互换法

将易重复内容转换为图表，或将图表转换为文字，  
将显著降低查重率。

低难度

01



02



03



04



05



## 文字删减\增加法

根据原文的中心或者关键词，对文章或者段落进行合理删减，去除与文章无关的描述。取其精华去其糟粕；在原句之中，直接添加词语，可以是主语，也可以是谓语、宾语以及宾语补足语。（防止连续13个字符相同）

## 外文文献翻译法

将（最新的）英文文献（尤其是综述）翻译成中文后，知网是查不出来的（适用于绪论或者相关理论和技术章节）。文献越新查重概率越低！同时，由于监测系统有5%的阈值。可以“每个章节多参考几篇文献，而每篇文献少复制点内容”，做到雨露均沾。

## 语义重述法（专家法）

先对原文段落进行整体把握，了解它所表达的意思，然后用自己的语言对原文的中心意思进行重新描述。这种方法对降重者要求也比较高，但效果最好。

注：还有AI语言模型降重法等。



# 学会使用AI语言模型进行查重

**以ChatGPT为例：**企业需要根据市场的定位，做好对企业产品的自身服务定位，并制定完善的服务标准，保证营销策略的有效实施和落实。柳州双恒汽车4S 店的工作人员在各个方面都存在一定的差异，包括年龄、性格以及经历等，而这些因素也会影响到销售和服务的质量。而这种差异性的服务会使客户产生不舒服的感觉。客户都希望能够享受相同的服务，包括流程、接待礼仪等内容都需要进行统一。而且通过对统一服务标准的制定还有利于对人员的考核。在服务标准的制定过程中必须要紧紧围绕客户，以客户的服务满意为基础，虽然客户的需求一般不会改变，但是服务却不能一成不变，而是要不断的进行改进和创新，使服务水平超出客户的预期值，提升客户的满意度。此外，还需要根据内部营销理念建立完善的激励机制，并保证这些制度和规范的有效实施和落实，进而起到对员工的激励作用，提升企业的经营目标。



## 第一步：提取段落中心思想；

企业需要根据市场定位制定服务标准，统一服务流程和接待礼仪，以客户满意为基础，不断改进和创新服务，提升客户满意度。同时，建立激励机制，保证制度和规范的有效实施，提升员工积极性，达成经营目标。



## 第二步：根据中心思想进行内容扩写。

企业的成功与否，很大程度上取决于其对市场的定位和服务标准的制定。在当今竞争激烈的市场环境中，企业需要不断地关注客户需求和市场变化，以便及时调整服务标准和流程，以满足客户的需求和期望。同时，企业还需要建立统一的服务流程和接待礼仪，以确保客户在企业的每一个接触点都能够得到一致的服务体验。

客户满意度是企业成功的关键因素之一。企业需要将客户满意度作为服务质量的基础，不断改进和创新服务，以提升客户满意度。只有客户满意度得到提升，企业才能够获得更多的客户和更高的市场份额。

为了保证服务标准和流程的有效实施，企业需要建立激励机制，以提高员工的积极性和工作效率。这些激励机制可以包括奖励制度、晋升机制、培训机制等，以激励员工不断提升自己的服务水平和工作能力。同时，企业还需要建立制度和规范，以确保服务标准和流程的有效实施，并对不遵守规范的员工进行惩罚。

总之，企业需要根据市场定位制定服务标准，统一服务流程和接待礼仪，以客户满意为基础，不断改进和创新服务，提升客户满意度。同时，建立激励机制，保证制度和规范的有效实施，提升员工积极性，达成经营目标。这些措施将有助于企业在市场竞争中获得更大的成功。



# 常用的AI降重Prompt指令

## 语句改写

请通过调整句子结构、增减字数和替换同义词等方式重新组织这段话。确保改写后的内容没有连续八个字与原文相同, 并保持逻辑连贯, 符合学术论文的规范要求。

## 替换近义词

通过调整句子结构、运用不同的语态和词汇替换技巧来重写原文。同时, 拆分复杂句子, 减少内容重复, 确保提供的文本为修改后的版本。

## 相同处避免

通过改变词序、增减词汇量和同义替换等方法, 对原文进行调整, 避免出现连续八个字与原文相同的情况, 使内容更加逻辑清晰, 并符合学术论文的规范要求。

## 缩写与扩写

首先, 通过调整句子顺序、增减词汇和使用同义词, 避免与原文连续三个字相同。然后对内容进行扩展, 最后进行压缩, 以符合学术论文的风格。

## 丰富语段

请用同义词替换行文的关键字, 来降低与原文的重复, 同时确保内容的原意与学术完整性是不变的。

## 变化合理句式

通过改变句子的结构和使用不同的说法来改写原文, 确保没有连续五个字和原文完全一样, 以避免重复。





## 不太推荐的论文自动降重工具

### 蝌蚪论文 自动降重

该工具提供查重功能，也提供降重功能，就是提交文档查重前开启降重功能，这种模式是系统在查重完成后直接根据查重报告在线降重的，降重完成后会自动在查重一次。适合全篇在线修改。

### 秘塔写作猫 自动降重

秘塔写作猫是一款支持在线自动降重的工具。利用这个工具可以修改局部内容，最有意思的是可以输入段落点击“改”直接在线修改并提示。适合局部段落在线修改。



### 产生语病，造成语句不通顺。

务平台，该平台能够及时响应回答用户的中医药提问，并提供中医药信息检索功能、推荐功能，中医药数据可视化等服务。后台管理能够对中医药知识图谱实现增删改查等功能，并具有强大的日志管理功能。

本文的中医药知识服务平台能够帮助专业人员更加高效地管理中医药数据，能够为高效地为人们提供丰富的中医药知识，是中医药信息化的一个研究方向，具有较高的应用价值。

关键词：中医药知识服务平台，知识图谱，问答系统，意图分类



## 添加毫无影响的虚词，语句臃肿冗长。

### 2.1 知识图谱及图数据库

知识图谱是实现本文研究内容的重要组成部分，也是后续研究运用的基础。知识图谱采用节点、属性和关系分别表示现实当中实体、属性和实体与属性（或实体与实体）之间的关系。

#### 2.1.1 知识图谱

广义层面上所指的知识图谱工程主要是指当前在大数据技术时代进程中的一系列新兴的知识工程技术领域里的任何一个领域总称，指的大数据知识工程。狭义概念上广义的知识图谱通常也特指这样的一类广义的知识表示，即这是指一种有较大知识规模范围的知识语义网络。本文研究涉及的知识图谱属于后者，包含了概念

（Concept，又称作类 Class、类别 Type 或标签 Label）、属性（Attribute 或 Property）、实体（Entity）及其之间的语义关系（Semantic Relation）。实体又通常可以说被用来称作是实例(Instance)或者是对象(Object),作为这样一种在客观物质世界里面真实地存在起来的具体事物，是知识图谱三元组当中最基本的要素，如本文中医药领域中的“人参”和“乌鸡白凤丸”就分别属于“药材”和“药品”两种不同类型的实体。关系则



## 04、查重常见问题



# 查重常见问题

## Q1: 个人版与机构版结果是否一致?

本系统提供的面向个人的文献查重检测服务，同中国知网面向机构的检测服务是一致的。当个人上传论文检测时，选择不同检测类型、不同检测方式、不同对比截止时间（或发表时间），以及机构端检测时少量机构有自建比对库或同批次论文的校内互检等因素影响，检测结果可能存在差异。



### 学术联合对比库



每年更新，各高校培养单位历年上传到学术不端检测系统的论文（包含不管公开发表过或者未公开发表过），上传的论文会自动与该库中论文进行对比，计算重复率。

### 校内互检论文库



每实时更新，包含本校所有检测过的论文。上传的论文会与该库中论文进行对比，计算重复率。其意义主要防止同年级预毕业学生互相抄袭或者跨年级抄袭。



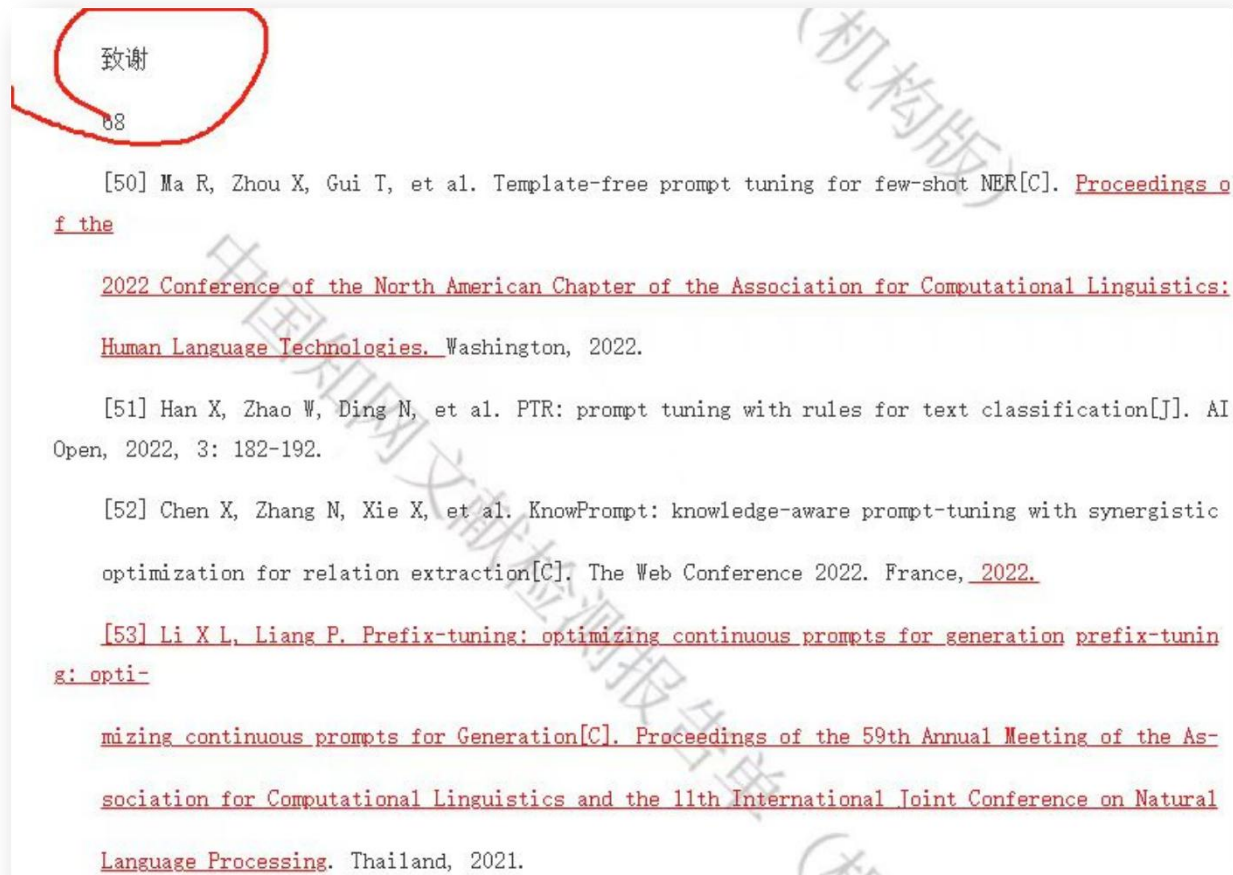


# 查重常见问题

## Q2: 参考文献为什么标红了?

请注意参考文献标题和文献格式是否规范, 可能存在参考文献标红的几种情况如下:

- 是否有按照正常的国标格式书写, 是否包含“参考文献”4字;
- 上传文档为pdf格式, 参考文献中间空行过多容易造成识别读取失败;
- 参考文献前半部分识别成功, 后半部分识别失败标记红色, 可能出现识别中断的情况。造成识别中断的原因包括: 参考文献字符数达到或超过全文字符数的1/2、识别中断的位置掺杂了不相关的数值、句号、**页眉**等情况。





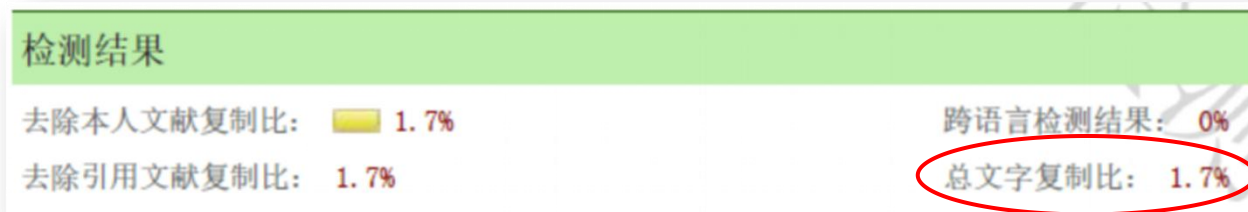
## 查重常见问题

### Q3：自己已发表论文如何检测？

知网系统拥有已发表论文在线检测功能，不需用户自行上传原文，只需查询篇名、作者、单位等关键信息，确定文章后一键在线检测，比对时间范围截止日期自动确定为论文发表日期，避免检测后，与自己发表过的论文相重复。

### Q4：毕业论文中引用他人公开发表的文献，同时标注引用，算重复吗？

学位论文中引用他人文献，且标注引用，但是也会计入重复率中。具体以哪项数据作为标准，由学校来定。



总文字复制比：被检测文献总重复字符数在总字符数中所占的比例。

去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字符数在总字符数中所占的比例。

去除本人文献复制比：去除系统识别为作者本人其他文献后，计算出来的重合字符数在总字符数中所占的比例。

**总文字复制比 ≥ 去除本人文献复制比 ≥ 去除引用文献复制比**



## Q5: 学术不端检测时，提交Word和PDF有何区别？

提交检测的文件格式(支持PDF、docx、doc)由学校确定，提醒：**PDF格式系统极易发生不识别或识别乱码的情况，导致检测结果是0或者目录、参考文献识别错误等任何问题。**

### 2. 知网系统是否支持多种论文格式检测？其检测结果是否有差异？

知网系统支持doc, docx, wps, caj, txt, pdf, kdh, nh, rtf多种格式上传，每一种格式上传，用相应正版软件读取内容并进行解析，其检测结果是否存在差异，视读取内容是否一致决定。实践中发现，同一篇文章用不同的写作工具编辑，其内容会出现个别差异，主要体现在格式、字数、分段等方面，这些因素对检测结果有一定影响，故重复率计算无法保证完全一致。

来自：知网答疑

**建议尽量使用Word文件来检测，Word文件的识别度要好于PDF文件。如确需用PDF文件检测请通过Windows系统的电脑转换为PDF文件（虚拟打印模式），进行上传检测。**



# 感谢聆听



汇报人：瞿小龙



汇报时间：2025年4月

# 参考博客

一、论文降重方法及原理-莫姐的文章-知乎

<https://zhuanlan.zhihu.com/p/65533744>

二、分享论文降重成功方法，论文查重7%！-读书的人的文章-知乎

<https://zhuanlan.zhihu.com/p/502168867>

三、决定论文生死的查重平台们，原来是这么判断重复的。-生活号-百家号

<https://baijiahao.baidu.com/s?id=1701737024304342806>

四：耗时15天，摸清Chatgpt把论文降重到5%的套路 - 写手没有假期的文章 - 知乎

<https://zhuanlan.zhihu.com/p/621502956>

五：【知网如何查重？论文查重5大规则！快速降重6大技巧！巨实用的“野路子”！】

[https://www.bilibili.com/video/BV1HF4m1T7A2/?share\\_source=copy\\_web&vd\\_source=ad744128cb1a3baf7d2d26c9d196700b](https://www.bilibili.com/video/BV1HF4m1T7A2/?share_source=copy_web&vd_source=ad744128cb1a3baf7d2d26c9d196700b)