

中文论文查重介绍及原理分析：为什么论文在知网会「标红一片」？

论文无意重复的六大「罪魁祸首」，就像埋在土里的地雷，稍不注意就会把人炸得「粉身碎骨」：



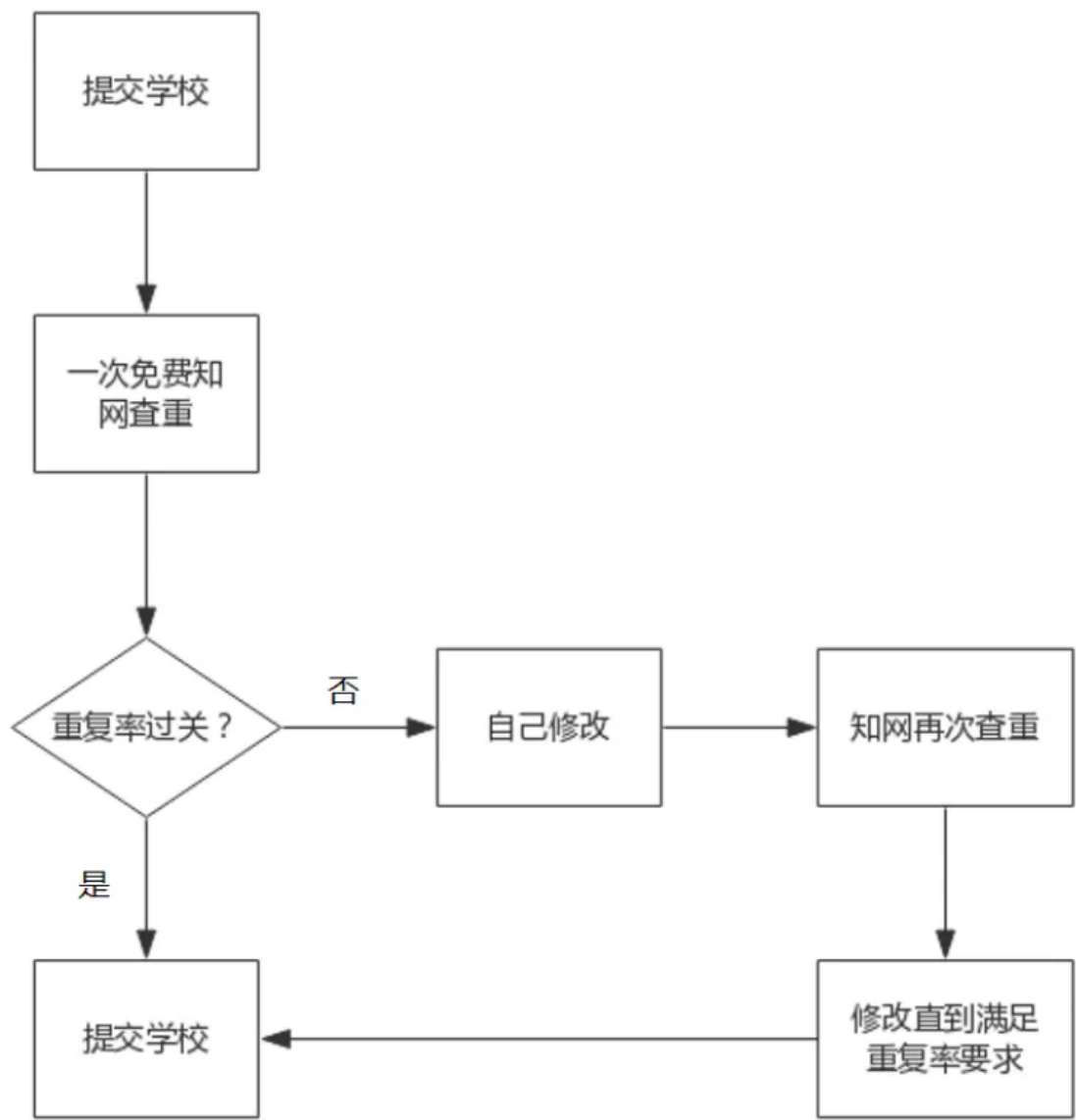
3-1 容易造成论文重复的潜在「地雷」

但是敌暗我明，我们在写作时该如何有效避开这些地雷，降低论文重复率呢？这就需要我们了解论文完成后查重的基本流程、学校老师和期刊编辑选用的查重软件以及它们背后蕴含的查重原理。

查重流程

查重的基本流程，一般有以下三种情况：

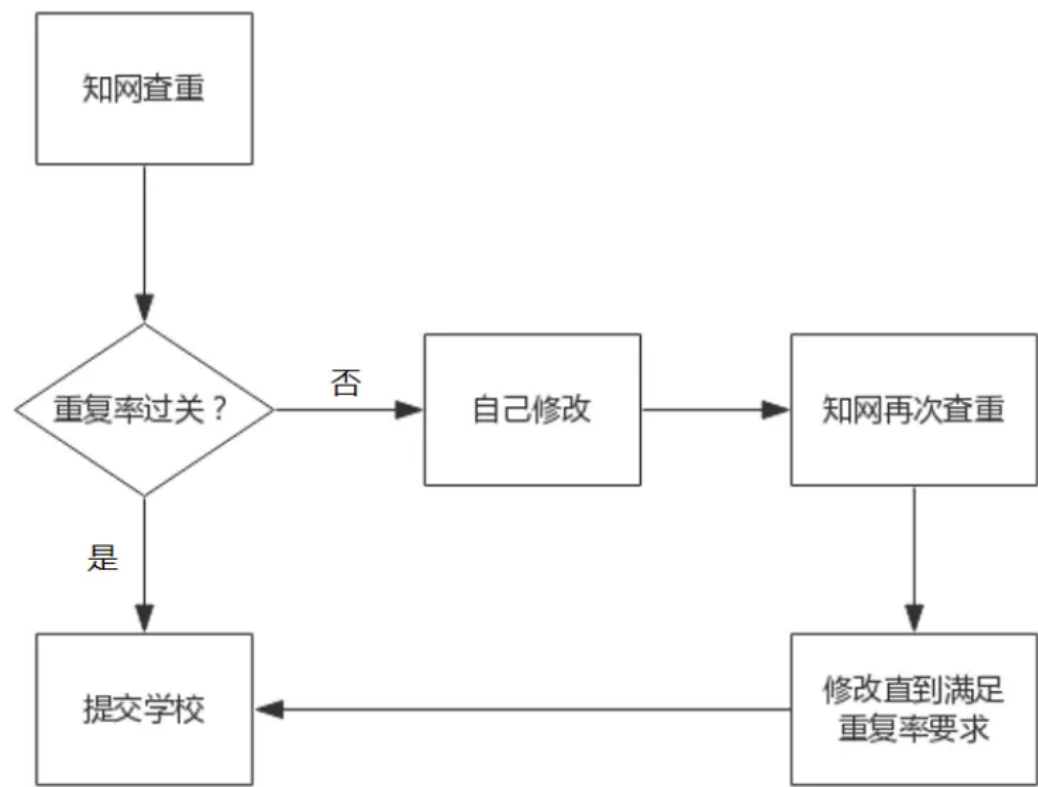
情况一：学校提供一次免费的毕业论文查重机会



加载中...

3-2 查重情况一

情况二：学校不提供免费的查重机会



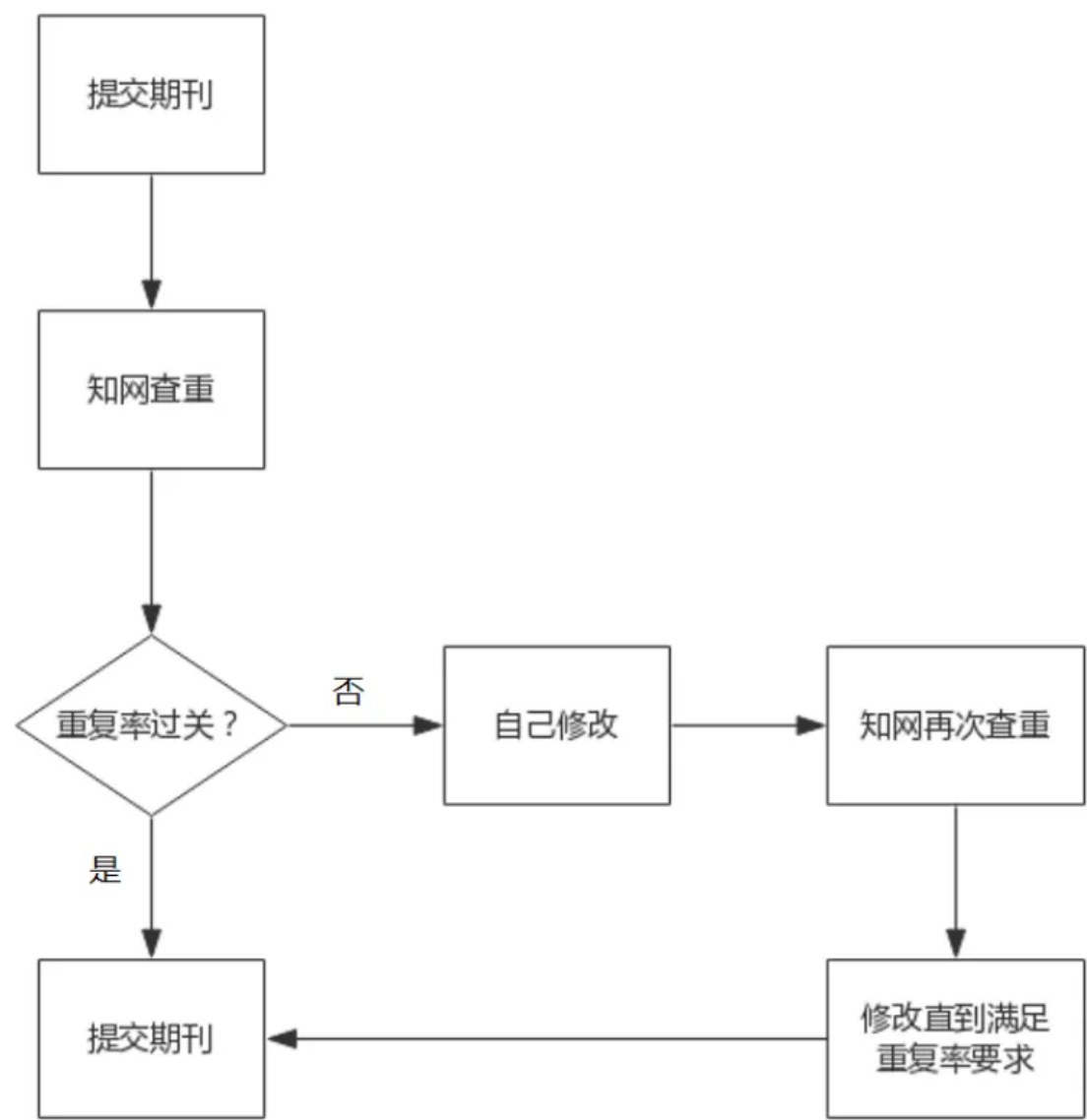
加载中...

3-3 查重情况二

在第一步时，有些同学贪图便宜，会使用其他软件进行查重，但是由于两者的数据库和查重原理不同，最后的结果和知网查重相去甚远，导致最后的重复率不满足要求，白白花了冤枉钱。

此外，免费的查重软件更需要警惕，天下没有免费的午餐，使用其他软件存在很大的**论文泄露**风险。

情况三：投稿中文期刊的论文



加载中...

3-4 查重情况三

需要注意的是，知网查重是不对个人用户开放的。真正有官方查重名额的，绝大部分是和知网有密切合作的第三方机构或者高校，除非个人在某宝上购买查询。

知网查重软件

从上述查重流程中，我们不难发现，知网查重具有举足轻重的地位。作为国内三大学术论文数据商（知网、万方、维普）之

一，在论文检测市场中，知网是绝大多数高校和中文期刊定稿论文必用的查重系统，搜索量最大，万方次之，维普最小。

越来越紧的学术政策下，高校毕业论文或者杂志社的投稿期刊文章，都要首先经过知网学术不端系统的检测，只有查重比例合格了才能进入下一环节。

从权威程度、知名度、市场口碑看，知网查重是名副其实的江湖老大哥。

知网是如何「查重」的

中国知网的 CNKI 工程是以实现全社会知识资源传播共享与增值利用为目标的信息化建设项目，从硕博研究生学术论文强制使用开始，几年的发展快速延伸到所有 211 和 985 院校，其他高校也开始以抽检的方式在陆续使用，使用范围包括出版社的审稿、职称评定、科研成果的查重等。目前，知网学术不端系统的数据库包含博士和硕士学位论文、期刊和会议论文、专利、网页数据、图书等文献资源。

起初，各个领域使用的知网数据库是一样的，后来随着社会发展和市场的需求，知网检测系统针对性越来越强，不同领域的数据库范围开始细分，使用对象不同，数据库也有所差别：

知网系统分类	适用人群	使用范围	数据库差别
大学生论文检测系统 PMLC	专科、本科生	学位论文	没有学术联合比对库
学位论文学术不端检测系统 TMLC2	硕博生	学位论文	没有大学生联合比对库
科技期刊学术不端检测系统 ALMC	投稿或评职称人士	科技期刊论文	没有大学生联合比对库和学术联合比对库
社科期刊学术不端检测系统 SMLC	投稿或评职称人士	社科期刊论文	没有大学生联合比对库和学术联合比对库

加载中...

知网学术不端系统分类及特征

知网查重的原理分析

关于知网的查重原理，网上众说纷纭，有谣传，有谬误，笔者搜集整理相关资料，并结合自己**近十年**的学术论文写作和论文查降重经验，概括了以下三条原理：

查重原理 1：句群语义模糊算法

知网论文查重采用最先进的语义模糊算法，关注**句子中心意思和关键词**，「了、着、的、可以」等虚词的修改并不会降低重复率。从图 3-5 可以看出，虽然查重报告左右两边红色部分内容并不完全一致，但是系统会智能判定为语句重复。

原文内容		相似内容来源
1	此处有 116 字相似 提高到77.54%，转鼓系数从53.48%提高到54.12%，平均粒度由23.42%增加至24.05%。冷风烧结时，由于料层自蓄热，下部料层热量过剩，温度较高，而上部料层热量不足，温度较低[]；同时，上部因抽入冷风急冷过快，烧结液相来不及结晶，形成玻璃质较多，且内应力和裂纹较大[117]。因此，表层烧结矿的强度较低。废气循环技术可使烧结用助燃气的物理热量显著提升，这部分热量有效补充了料层中上部热量不足，缓和了中上部烧结矿的急速冷却速度，使矿物熔融结晶发育条件变好(熔融	攀钢高炉高产实践 范云东,王戈,李劲明-《中国冶金》-2003-08-15 (是否引证：否) 1. ,改善了料层过湿现象,提高了料层的透气性。③采用热风烧结。在烧结中国冶金生产过程中由于布料偏析和自动蓄热作用,造成上下部温度不均,同时,上部因抽入冷风急剧冷却,使烧结液相来不及结晶,形成玻璃质,并产生较大的内应力和裂纹,降低了表层烧结矿的强度。热风烧结以热风的物理热代替部分固体燃料的燃烧热,使烧结料层上下部热量和温度的分布趋向均匀,热风烧结具有改善表层烧结矿强度的重要作用。此外,由 刘政群-论文-20120518 论文-《学术论文联合比对库》-2012-05-25 (是否引证：否) 1. %左右,有待于进一步组织工业试验。热风利用技术研究进展 热风烧结 在烧结生产中由于布料偏析和自动蓄热作用,料层下部热量过剩温度较高,而料层上部热量不足,温度较低。同时,上部因抽入冷风急剧冷却,使烧结矿液相来不及结晶,形成大量玻璃质并产生较大的内应力和裂纹。因此降低了表层烧结矿的强度[34,35]。热风烧结是在点火之后,烧结机首部三分之一长度上,往料面喷入温度为250℃~1000℃的热空气或热废气。
	此处有 42 字相似 融结晶更充分)、热应力引起的裂纹减少(矿物结构改善),因而改善表层烧结矿的强度。同时,这些热量将使料层内部热量和温度分布趋于均匀,可减少形成脆性、薄壁、大孔结构的可能性,提高烧结矿的整体强度[],降低烧结矿中FeO的质量分数,显著提升烧结矿的还原性[]。图3-10(a)所示为对比了循环气体温度的影响,对比可知:(1)	攀钢高炉高产实践 范云东,王戈,李劲明-《中国冶金》-2003-08-15 (是否引证：否) 1. 结料层上下部热量和温度的分布趋向均匀,热风烧结具有改善表层烧结矿强度的重要作用。此外,由于配料中国体燃料减少,固定碳分布趋于均匀,减少了形成脆性、薄壁、大孔结构的可能性,有利于烧结矿强度的提高,采用热风烧结以后,外部供热量增加,降低了固体燃料消耗,两个阶段混合料固定碳平均下降0.13%,折合成无烟煤则固体燃料消耗 莱钢3×265m~2烧结机烧结矿余热充分利用的实践 赵红光,李兴义,元玉辉,胡守忠,毛友庄;-《冶金能源》-2009-07-30 (是否引证：否) 1. ·1热风烧结可行性分析热风烧结使烧结料层上层温度升高,降低了上下部料层的温差,有利于减轻因上下层温差大引起的烧结矿质量不均匀,减少了形成脆性、薄壁、大孔结构的可能性,提高了烧结矿转鼓强度。环冷机烧结矿

加载中...

3-5 知网查重报告

其他片段：

酝酿已久的个人存款账户实名制度终于颁布实施，这表明我国个人存款管理制度日趋向国际惯例靠拢。个人到金融机构办理存款时需要出示个人身份证。

疑似文章片段：

- 个人存款实名制度正式实施。

- 接近国际惯例的个人存款制度开始实施。

分析：

被检测片段的核心意思实际上是个人存款实名制实施。如果它检测到其他的文献中有类似意思的片段会认为这可能是重复的，它的原理其实就是检测句子的中心意思和关键词，如果两者都类似就会被判为重复。

查重原理 2：划分章节分别检测，阈值 5%

整篇论文上传后，系统会自动根据文章生成的目录检测该论文的章节信息，然后系统会将论文**分章节检测**，可以获得每一单章节的复制比，同时目录显灰色不参与正文检测；否则会自动分段按照 1 万字符左右检测，同时，目录有可能被当成正文检测，重复就会标红。

中国知网对该套查重系统的灵敏度设置了一个阈值，该**阈值为 5%**，以段落计，低于 5% 的抄袭或引用是检测不出来的，这种情况常见于大段落中的小句或者小概念。

分章节片段：

第一章

盼望着，盼望着，东风来了，春天的脚步近了。

第二章

一切都像刚睡醒的样子，欣欣然张开了眼。山朗润起来了，水涨起来了，太阳的脸红起来了。小草偷偷地从土里钻出来，嫩嫩的，绿绿的。园子里，田野里，瞧去，一大片一大片满是的。坐着，躺着，打两个滚，踢几脚球，赛几趟跑，捉几回迷藏。风轻悄悄的，草软绵绵的。

第三章

桃树、杏树、梨树，你不让我，我不让你，都开满了花赶趟儿。红的像火，粉的像霞，白的像雪。花里带着甜味儿；闭了眼，树上仿佛已经满是桃儿、杏儿、梨儿。

第四章

花下成千成百的蜜蜂嗡嗡地闹着，大小的蝴蝶飞来飞去。野花遍地是：杂样儿，有名字的，没名字的，散在草丛里，像眼睛，像星星，还眨呀眨的。

未分章节片段：

盼望着，盼望着，东风来了，春天的脚步近了。

一切都像刚睡醒的样子，欣欣然张开了眼。|山朗润起来了，水涨起来了，太阳的脸红起来了。小草偷偷地从土里钻出来，嫩嫩的，绿绿的。园子里，田野里，瞧去，一大片一大片满是的。坐着，躺着，打两个滚，踢几脚球，赛几趟跑，捉几回迷藏。 |风轻悄悄的，草软绵绵的。

桃树、杏树、梨树，你不让我，我不让你，都开满了花赶趟儿。红的像火，粉的像霞，白的像雪。花里带着甜味儿；闭了眼，树上仿佛已经满是桃儿、杏儿、梨儿。

花下成千成百的蜜蜂嗡嗡地闹着，大小的蝴蝶飞来飞去。野花遍地是：杂样儿，有名字的，没名字的，散在草丛里，像眼睛，像星星，还眨呀眨的。

分析：

从上面两个例子可以看出，如果在文中有标记章节，那么系统会分章节查询，也就是按照第一章、第二章等分别检测，然后每章节中发现重复率大于 5% 的阈值就会列出，否则不会出现在重复列表中。这是主动分章节的情况。

但是遇到没有分章节的大段文字，查重系统也会帮忙分章节然后再查重，然后你会发现同一文章的多次查重结果是有轻微差异的，但是这个差异不会太大，而**做到合理的分段**有时候也会降低我们的重复率。

查重原理 3：疑似部分二次检测

查重系统最厉害的地方就是这个功能，即它会对疑似的部分进行**二次检测**。

以下面这段文字为例：

我曾经看过一段文字，它是这么说的：「生而为人，我很抱歉」。我认为这有点悲观。

「生而为人，我很抱歉」这句话是很容易被发现重复的，如果第一次被系统判定为疑似重复，那么系统就会加强这句话和相关的讨论太宰治作品、影视作品的文字或者网络文章进行对比，并且前后文也会加强检测。经过如此的重重检测，想逃避重复的事实难如登天。

幸运的是，有的内容还没有被系统纳入查重的范围，它们分别是

(1) 公式，比如数学公式就不会被查重

如：

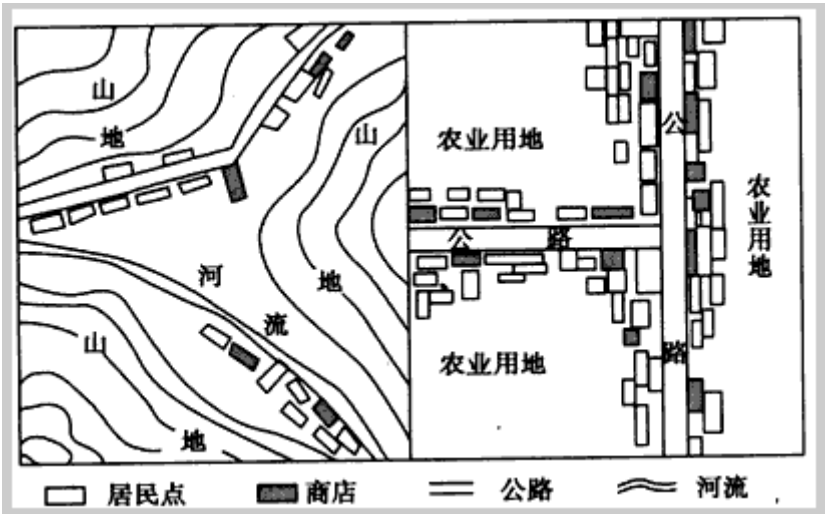
$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

加载中...

3-6 数学公式

(2) 图形和其中的文字也是没有参与查重的

如：



加载中...

3-7 图片及图例

但是随着技术的发展，光学识别技术是可以实现图片查重的，所以，很有可能有一天查重系统升级导致图片也被查重，因此大家不要掉以轻心。

(3) word 文件中的域代码，比如 Endnote 在论文中生成的域代码是没有查重的

{ BIBLIOGRAPHY [可选开关] }

加载中...

3-8 Endnote 域代码

(4) MathType 打印的公式符号也是安全的
如：



加载中...

3-9 MathType 公式符号

有了原理就能找到针对性的改写方法，下一节将介绍具体的降重改写方法步骤。

参考文献

[1] 2019 论文查重市场品牌分析[EB/OL].

<https://zhuanlan.zhihu.com/p/61425619>.

[2] 论文查重品牌介绍[EB/OL].

<https://zhuanlan.zhihu.com/p/66951072>.

[3] 中国知识资源总库系列产品[EB/OL].

<http://gb.oversea.cnki.net/kns55/default.aspx>.

[4] 揭秘：知网论文查重的规则及检测原理[EB/OL].

<http://www.cnkis.net/html/1095371058.html>. I

浏览器扩展 Circle 阅读模式排版，版权归 www.zhihu.com 所有