

了解知网查重原理，重复率 3% 以下其实很简单

【题记】对于论文原创度的合格标准，各个学校之间也不尽相同，有些学校定的标准线稍为宽松一些，重复率在 30% 或 20% 以下就算合格；而有些学校则比较严格，规定要在 10% 甚至 5% 以下才算合格。5% 的要求真的是极度严格了，因为常用字就这么三千多个，专业领域的词汇也都是通用的，随便写点什么都可以查找到相似的句段……

国内应用最广泛，最权威的是知网系的检测系统，也是目前高校普遍采用的系统，关于知网检测的方法和步骤，主页君在 3 月 24 日的文章“论文定稿后查重，用知网分解”（点击进入）进行了详细的介绍，今天，白熊以知网检测系统来作为案例来做一些具体的分析。

1. 在知网查重报告中，标黄色的文字代表这段话被判断为“引用”，标红色的文字代表这段话被判断为“涉嫌剽窃”。

2. 在知网查重进行中，检测系统只能识别文字部分，论文中的图片、word 域代码、mathtype 编辑的公式、是不检测的，因为检测系统尚无法识别这些复杂的内容格式。你可以通过[全选]——[复制]——[选择性粘贴]——[只保留文字]这样步骤的操作来查看具体的查重部分。另外，在编辑公式时，建议使用用 mathtype，不要用 word 自带的公式编辑器。

3. 论文中的表格内容数据是可以识别的。如果表格的内容有很大的重复度，那么可以把表格截图保存，再放到论文中去。

4. 论文中引用的参考文献部分也是会计算相似度的，所以引用的时候，最好的方式就是先理解内容，再用自己的话写出来。

5. 知网检测系统对论文的分节是以“章”作为判断分节的。封面、摘要、绪论、第一章、第二章、等等这样一系列的都会各自分成一个片段来检测，每一个片段都计算出一个相似度，再通过这样每章的相似度来计算出整篇论文的总重复率。

6. 有一点可能很多人都已经知道了，就是在知网查重系统中，是以“连续多个个字重复”做为识别标准。如果你能够通过修改，让你的论文任意一句话都找不到连续多字与别人的文章相同，这样就检测不到。

7. 当知网查重系统识别到你论文中有某句话涉嫌抄袭的时候，它就会对这句话的前面后面部分都重点进行模糊识别，这个时候判断标准就变得更严格，仅仅加一些副词或虚词（比如“的”、“然后”、“但是”、“所以”此类词语）是能够识别出来的。但这样子的识别方式有时候会显得不是很智能，可能会扯上一篇完全不相干的论文说是内容相似，这个就属于比较无奈的情况了。

8. 知网的对比文库里不包括书籍，教材等。但有一个问题要注意，当你“参考”这些书籍教材中的一些经典内容时，很可能别人已经“参考”过了，如果出现这样子的情况，那就会被检测到相似。有些同学会说，那我用自己的话去重新写一下，这样就不会被检测到啦。这样的方法，理论上是对的，但实际上是，这些经典的句段，已经有无数的人引用过了，也已经被无数的人通过各种改写的方式引用到论文中去的，所以，要写出一段跟别人不一样的话，还真

是有点难度，这个也要拼点运气。

9.网络上的某些内容也是在知网的数据库里的。比如：“百度文库”、“道客巴巴”、“豆丁网”、“互动百科”、“百度百科”。作者查重的时候，甚至还遇到很多奇葩的网站，神马“东方财富网博客”、“人大经济论坛”。所以，选择网上的内容时要慎重。大家在网上搜索资料的时候，也不要直接引用，知网查

重系统在检测文章的时候，也会进行网络搜索的，特别是一些网络学术资料比较多的一些网站，如百度文库、百度知道、360 文库等这类网站上面，很容易搜索到相似的信息。

10.大家可以多参考一些国外的资料，因为在知网的对比文库中，外文资料相对比较少一些，而且用自己的理解翻译成中文的话，几乎不会检测到重复。

11.各位在论文提交到学校之前，一定要自己先到网站查一下，如果有检测出来相似度较高的片段，自己先改一改。论文修改一次以后，不要以为就肯定能过了，因为知网系统会根据论文内容的不同自动调整着重检测的段落，所以有时候第一次查重的时候是正常的，一模一样的句子，第二次检测的时候会判断为“抄袭”。这也是没有办法的，只能多查多改。

12.知网官方检测系统不对个人开放，学生自己是无法自行到知网去检测论文的，只能通过第三方检测平台进行，除了知网之外，国内应用比较广泛的论文查重系统还有“万方数据”、“checkpass”、“paperpass 检测”，“paperrater 检测”这几个。总的来说，万方是最松的，paperrater 这两个最严。checkpass 检测算法最先进，价格适中，从费用方面来说，知网的价格是最高的，其它的相对便宜一些，但是检测结果会与知网有区别。所以一般建议先用便宜的系统查一下，把能改的都改掉，再用知网系统去查，这样子比较划算。在这里，还是推荐大家使用 checkpas 检测系统，检测算法先进，结果准确，最为接近知网，非常适合论文初期、中期以及定稿后初次检测。