现在高校对于硕士和博士论文采用的检测系统，是由知网开发的。但该软件的具体算法，判定标准，以前一直不清楚，  
  
本文是从知网内部工作人员哪里拿到的，揭示了知网反抄袭检测系统的算法，如何判定论文是抄袭，以及如何修改来通过的秘籍。发出来造福大家。

引用：  
  
1、对格式的要求  
    知网学位论文检测为整篇上传，格式对检测结果可能会造成影响，需要将最终交稿格式提交检测，将影响降到最小，此影响为几十字的小段可能检测不出。都不会影响通过。系统的算法比较复杂，每次修改论文后再测可能会有第一次没测出的小段抄袭（经2 年实践经验证明，该小段不会超过200 字，并且二次修  
改后论文一般会大大降低抄袭率）  
  
2、对比库  
    对比库为：中国学术期刊网络出版总库,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,个人比对库,其他比对库, 部分书籍不在知网库，检测不出抄袭。知网库是国家指定的论文检测对比库，国家指定高校论文检测系统为知网学位论文检测系统，该系统是目前效果最好、范围最广的官方检测系统，所有高校都是知网的检测系统，这是教育部出于全国学术不端公平性考虑而实施的。  
  
3、关于分段和分章出结果  
    上传论文后，系统会自动检测该论文的章节信息，如果你学校的目录设置符合知网系统内置的分章判断条件，系统就会按章检测，分章出结果，否则会分段出结果。关于分段或分章主要涉及4 中的阀值。诚信论文提醒，不论是分章还是分段，保持和学校一致即可。  
  
4、引用的能检测出来吗？  
    有的同学问：“我明明引用了别人的段落或句子，为什么没有检测出来？”也有的同学问：“我的引用标注了出处，为什么还算抄袭？”首先，引用算不算抄袭，与标注出处没有任何关系，引用能不能检测出来，与系统准不准确也没有关系。所有这些都靠系统的阀值来决定。中国知网对该套检测系统的灵敏度设置了一个阀值，该阀值为3%，以段落（或章节）的字数来计算，单篇文献低于3%的抄袭或引用是检测不出来的，这种情况常见于大段文字中的小句或者小概念。举个例子：假如检测段落1（第一章）有10000 字，那么引用A 文献300字（10000 乘以3%=300）以内，是不会被检测出来的。若引用B文献超过300字，那么B文献分布于第一章中的抄袭都会被红字标注，不管位于第一章何处，即使打断成句子，只要超过20字就会被标注。①实际上这里也告诉同学们一个修改的方法，就是对段落抄袭千万不要选一篇文章来引用，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。②关于一些同学问引用的为什么也算抄袭，这里主要是因为知网的阀值问题，高于3%的统一算抄袭，也就是说引用于抄袭的临界就在3%之间。一旦你超标，即使你标注了引用也无济于事。所以请同学们注意。我们举例说明：某篇论文第一章有5000 字，那么第一章中，我们就只能引用A文献150 字以下，否则会被系统认为是抄袭。第二章4000 字，那么我们只能引用A 文献120字以下，否则会被系统认为是抄袭。第三章8000 字，第四章7000 字，分别为240 字以下和210 字以下，以此类推。综上所述，引用超标的计算方式是按章计算，这与抄袭的计算方式是一样的。  
  
5、系统对一句话怎么才算抄袭？  
    一篇论文的抄袭怎么才会被检测出来？知网论文检测的条件是20字单位以上的相似或抄袭都会被红字标注，但是必须满足4里面的前提条件：即你所引用或抄袭的A 文献文字总和在你的各个检测段落（各章）中要达到3%。  
引用：  
  
6、抄袭的修改方式  
  
    针对标红文字的修改方式除了3中提到的外，还有改词、换句、改变描述方式（变原句为倒装句、被动句、主动句等）、打乱段落顺序、删除关键词汇、关键句等。经过实践证明，使用以上方法结合，可有效降低复制比，保证顺利通过。  
  
总体来说，我们需要在保证修改后句子通顺的前提下，尽量和原句在字面上保持不同。  
  
例1：例如下句：  
    过热故障中的过热与变压器正常运行下的发热是有区别的，正常运行时的其发热源来自于绕组和铁芯，即铜损和铁损，而变压器过热故障是由于受到有效热应力而造成的绝缘加速劣化，它具有中等水平的能量密度。  
  
几乎被标红，说明与相似文献存在重合和高度相似，经过以上方式结合，本句可改为：  
  
    过热故障中出现的过热容易与变压器正常运行下的发热相混淆，后者是因为其绕组和铁芯会出现铜损和铁损的现象，这是正常运行过程中的发热，而变压器过热故障是受到有效热应力造成的绝缘加速劣化，   
  
① 这里所指的300 字是一个大概值，并非临界值。引用的数量越低，就越不容易被检测出来。  
  
② 更新以后的CNKI 学术不端检测系统将这一阀值调整到了3%，以前是5%，意味着检测系统对引用的要求更加严格，但运用我们后面提到的方法也不是很难。具有中等水平的能力密度。  
  
这样修改几乎可以降低抄袭率一半。  
  
例2：在看下面一个例句：  
  
3.7.1.2 在透明水杯的清水中放入少量纤维进行搅动，便可以直观地发现纤维呈立体悬浮状乱向分散，且长时间放置都不会有太大变化，说明合成纤维的质量较好；质量差的纤维经搅动后可能分散，但时隔不久便会上浮为一絮状层。质量差的纤维在混凝土的实际配制过程中多不易均匀分散。   
  
本段完全被标红，修改方式只有一种，就是打乱顺序，重新组织。  
  
3.7.1.2 将少量纤维放入盛装清水的透明容器中，边搅动边观察纤维变化情况，如果合成纤维质量较好，那么就可以直观地看到纤维呈立体悬浮状分散，随着时间的推移，位置也不会发生明显的变化；若合成纤维质量较差，那么搅动的过程中，纤维可能分散，并且容易上浮形成絮状层。质量差的纤维在混凝土的实际配制过程中多不易均匀分散。   
  
例3：下句：  
对施工单位或业主提出的设计变更要求要进行统筹考虑，确定其必要性，同时将设计变更对施工工期和费用的影响进行全面分析，非改不可的要调整施工计划，以尽可能减少对工程的不利影响。   
  
修改为：  
  
  
施工单位或业主一旦提出设计变更要求，要进行统筹考虑，考察变更的必要性，同时，将设计变更对施工工期、费用等方面可能造成的影响进行全面而科学的分析，遇到非改不可的变更要调整施工计划，尽可能将其对工程的不利影响降到最小