

怎么降低论文查重率?

论文自己写，就能高枕无忧了吗？

论文重复是一件祸患无穷的事，虽然在当前不完善的学术不端处理制度下仍存在不少漏网之鱼，但是铤而走险终究不是正道。

其实，在具体写作论文的过程中，有很多环节容易造成论文重复需要格外注意的。

重复可以分为故意重复（即直接挪用他人写作成果）和无意重复（即自己写的内容也会重复，属于非故意造成的重复）。

故意重复，比如直接复制粘贴造成雷同，没有分析的必要，不在讨论之列。

而分析无意重复的根源，则需要从论文的关键内容出发，比如题目、研究背景、研究进展等部分，找到造成中英文论文无意重复的六大「罪魁祸首」。

1、题目的核心要素类似，容易重复

由于题目中包含的核心关键词在同一领域很可能一样，因此如果研究内容和已发表论文的内容相似，那么就很有可能出现题目高度重复的情况，如表 2-1 所示：

中文题目 1	再生混凝土耐久性研究现状综述
中文题目 2	海砂混凝土耐久性研究现状综述
英文题目 1	Material properties of cold-formed high strength steel at elevated temperatures
英文题目 2	Mechanical properties of cold-formed steel tubular sections at elevated temperatures

表 2.1 高度重复的论文题目（中英文各一例）

对比中文题目 1 和 2，可以发现两篇综述的题目只有两个字不同，题目重复率高达 85.7%（重复字数 12 字/总字数 14 字），研究对象高度相似，一个是「再生混凝土」，一个是「海砂混凝土」，且研究变量都是关于混凝土的耐久性。

对比英文题目 1 和 2，可以发现两篇论文研究的对象相似（cold-formed high strength steel，高强冷弯型钢；cold-formed steel tubular sections，冷弯钢管截面），且测试条件相同（都是 elevated temperatures，高温），只是把研究对象的属性从材料属性（material properties）变成了力学属性（mechanical properties），可见两个题目结构相同且关键句子核心关键词高度相似，最终导致了较大程度的重复。

题目重复的主要原因是题目的核心要素相似，如研究对象、研究变量、研究条件、研究理论/方法/关系、范围等。这些核心要素常常被写在一些常用的表达中，比如题目的通常写法有（下划线表示研究对象、条件或方法、范围）：

- 成都 379 例新型冠状病毒肺炎的临床分析
- 高密度电法在废弃矿渣堆体底界面探测中的应用研究
- 护理干预对慢性乙肝患者睡眠质量的影响
- 基于深度卷积神经网络的地震数据断层识别方法

- 基于模糊层次分析的水稻化肥减施效果评估研究——以湖北省为例

虽然论文题目的字数很少，即使被查出有重复也不会对论文整体的重复率有较大影响，但是如果被论文答辩老师、期刊编辑发现高度重复，就很有可能判定我们的论文在重复前人的研究主题，创新性很差，极大概率被答辩老师一票否决或者被编辑拒稿。

同时，这也提醒我们，在提交毕业论文或者投稿前，先去搜索一下是否有高度相似的题目，如果有则需要改动。

2、研究背景类似，模仿后容易重复

在引言的写作中，首先要在第一段介绍研究背景，这在两种情况下容易造成重复：

情况 1：在引言的第一段中描述和国内、国际同行相似的研究背景。这时候，如果去模仿已经发表出来的同行论文，而且没有正确地改写，就容易造成重复。

原文 1：

近些年来，学位论文的抄袭剽窃、重复发表、低水平重复等道德失范的现象愈演愈烈，从一个侧面反映出高校教学质量和人才素质培养的缺陷问题。这些问题不仅影响教育教学质量的提高，而且严重影响了学术声誉、阻碍学术创新和学术进步，特别是研究型、知识型人才的培养。

原文主要描述了当前道德失范现象严重，影响恶劣且广泛。以此为研究背景模仿时，由于中文查重软件基于「句群语义模糊算法」（关于中文论文查重原理将在下一节中介绍），稍不注意，改写后的部分内容就会和原文比较类似。

重复案例：

道德失范现象近几年愈演愈烈，学位论文抄袭剽窃、重复发表甚至出现低水平重复。这从侧面折射出两个缺陷问题——高校教学质量不过关，人才素质培养有死角。这些问题的存在不利于提高教育教学质量、推动学术创新和进步，并且对学术声誉有极大的影响。

句子中划线的部分都是和原文雷同的地方，成片的重复说明这是一次失败的改写。

原文 2：

The transport properties of cementitious materials have been studied for many decades. This is because movement of aggressive species such as chlorides, carbon dioxide, oxygen, sulphates and alkalis are responsible for most deterioration processes affecting concrete structures including reinforcement corrosion, sulphate attack and alkali-aggregate reaction.

原文的中心意思是说明水泥基材料的传输性能很重要以及重要的原因。小领域同行很可能也需要表达类似的意思，如果想模仿并改写，但是把握不好的话，就会造成重复。

重复案例:

The transport properties of cement-based materials have been extensively studied. It has been found that the movement of aggressive species in cementitious materials such as chlorides, CO₂, oxygen, sulphates and alkalis could cause most deterioration processes which affect concrete structures such as reinforcement corrosion, sulphate attack and alkali-aggregate reaction.

模仿后的段落足足有 63.3% 的重复率（重复单词 31 字/总字数 49 字），其重复的单词和短语用下划线表示。

情况 2：某一个大课题成果较为丰富，可以就该成果连续发好几篇论文，而每篇论文的大背景大同小异，这样引言中的研究背景部分就容易造成自我重复。

比如，基于一个博士课题成果，往往可以发 3-5 篇 SCI 论文，由于是同一个课题，因此总的研究背景是相似的，这时候就容易造成引言第一段的重复。

3、综述当前研究成果

在引言中，容易出现重复的第二个部分是对前人研究成果的总结概括。如果没有高度概括前人的研究成果，就容易出现短语结构和关键词重复的情况，比如：

句 1：

要使房地产业健康发展，一方面要有完善的市场机制，另一方

面要有有力的财政政策和金融政策调控。

句 2:

房地产业的健康发展离不开完善的市场机制、有力的财政政策和金融政策调控。

虽然两句话分别用了关联词（一方面，另一方面）和并列结构，但是整体句子结构还是高度相似，而且「健康发展」、「完善的市场机制」、「有力的财政政策」以及「金融政策调控」都是雷同表达，所以出现了重复。

句 3:

We particularly consider the problem of capillary infiltration of saturated solution into a porous medium, which has been commonly used by a number of experimental and numerical studies on crystallization of salts in geomaterials.

句 4:

There have also been a number of experimental and numerical studies on the fire resistance of concrete materials.

短语 a number of experimental and numerical studies on（在某方面有许多试验和数值模拟研究）经常被用在引言中表示某个热点话题的关注度，因此就容易造成重复。

此外，在报道某种情况或数据时，由于数据和报道的对象都是统一的，就容易被判断为雷同，比如：

句 5:

2008 年 5 月 12 日，四川省汶川县映秀镇附近发生 8.0 级强地震，造成 69227 人死亡，374643 人受伤，17923 人失踪。

句 6:

发生于 2008 年 5 月 12 日四川省的「汶川大地震」我们还记忆犹新。8.0 级强地震夺走了 69227 人的生命，让 374643 民众受伤，还有 17923 人下落不明。

因为关于「汶川大地震」的发生时间、震级和带来的影响，都是十分确定的，所以在描述该类对象时，改写不到位就会出现重复。

4、相同或相似研究方法或数据，易重复

如果我们的论文研究方法和同行类似，往往需要参考已发表论文中的方法和具体步骤。其中那些较为标准和通用的方法步骤往往可以被借鉴。如果我们没做好改写，就容易造成语句重复。

同时，如果我们已经发表过使用某个试验方法或数值分析方法的论文，在新的论文中需要再次描述同一方法，就很容易造成自我重复。比如，我们有可能把已经发表的论文的方法部分作为一个模板，保持主体结构不变，只改动关键变量，比如材料名称、测试条件等。

主体结构不变，只改动关键变量造成的重复

论文 1:

试验试件严格按照 Concrete and reinforced concrete sewer pipes (GB/T 11836-2009) 选取, 其内径为 $r = 400 \text{ mm}$, 壁厚为 $t = 50 \text{ mm}$, 单节长度为 $l = 2.0 \text{ m}$, 接头方式为承插式, 以柔性橡胶圈形成密封。

论文 2:

我们严格按照 Concrete and reinforced concrete sewer pipes (GB/T 11836-2009) 标准选取试验试件, 制备了内径 r 为 400 mm , 壁厚 t 为 50 mm , 单节长度 l 为 1.5 m 的试件。其接头方式为承插式。我们用柔性橡胶圈进行密封处理。

仔细对比以上两段, 我们发现除了单节长度 l 略有不同之外, 其他部分几乎完全相同。这样的「套用」, 是通不过中文查重软件的筛查的。

再比如, 同一个课题组中的科研工作者都使用同一实验室、仪器及实验方法, 这时如果没有正确改写, 就容易造成论文的自我重复或与同门/同行重复。

而如果研究数据和同行相同, 也会造成数据重复, 请看下面真实的审稿意见:

「Checking through your manuscript, it is noted that large part of the data is drawn from already published papers. Since the present manuscript does not follow ethical practices in publishing, the journal cannot consider the submission for publication. Our journal is a high-ranking

scientific journal which strictly follows ethical practices.」

（「我们检查你的手稿后发现你的大部分数据来自已经发表的论文。由于本手稿在出版过程中不遵循伦理规范，本期刊不能考虑投稿发表。本期刊是严格遵守伦理规范的高水平期刊。」）

虽然，这些通用的试验方法很难有新的写法，作为作者，我们也知道自己研究的重点不在试验方法上，即便试验方法相同，研究的创新点还是存在的。可是，如果作者在提交论文之后无法通过计算机查重算法的初次审核，文章的重复问题就会上升到道德层次，将无法给编辑发现文章精华的机会，更有甚者可能使作者丧失再次投稿的机会。

5、研究变量相似，结果易重复

如果使用和已发表的论文中相同的试验或模拟方法，特别是和自己已发表论文中的实验方法相同，且测量的变量也相同，就会出现论文结果部分的表达与前人或自己论文重复。

句 1:

由图 20 可见，对于水平荷载作用下产生的水平位移，虽然孔洞的深度对于管道位移动力响应有显著的影响，但影响主要集中在频率段 0.1 范围内。在低频段 ($\omega < 0.1$) 内，孔洞对于管道的位移动力响应影响很小。

句 2:

从图 18 我们发现，虽然管道位移动力响应受孔洞深度的显著

影响,但是对于水平荷载作用产生的水平位移,其影响范围主要集中在频率段 0.1 内。它在低频段 ($\omega < 0.1$) 内影响很小。

通过对比上述两句,我们发现关于论文结果的论述,基本一致。虽然调换了语序,改变了图片编号,但是还是逃不过计算机查重的「火眼金睛」。

句 3:

The material properties obtained from the stress-strain curves from transient state test results under room temperature are summarized in Table 5, using the definitions of symbols as shown in Fig. 4.

句 4:

Under room temperature, the material properties obtained from the stress-strain curves from transient state test results are summarized in Table 2, using the curves as shown in Fig. 7.

第二句中相对于第一句的改动主要是替换了图表的标号,但都是同一个试验(transient state test,瞬态试验),且都是在相同测量条件下(under room temperature,室温下)测试相同变量(stress-strain curve,应力应变曲线)。

6、无法和难以避免的重复

目前的计算机查重算法将有固定表达的作者单位信息(Affiliations)、变量说明(Notations)、相对固定的致谢(Acknowledgement)、利益冲突声明(Conflict of Interest

Statement) 和道德申明 (Ethical Statement) 都纳入查重范围, 因为这五部分一般都会被算作重复部分。不过答辩老师、期刊编辑大部分都理解这些部分的重复, 而且它们所占的篇幅很小, 因此不会对全文的重复率有大的影响。

以医学中的某标准道德申明为例:

Ethical Statement:

Written informed consent was obtained from all participants before the commencement of the study.

(在研究开始前, 我们均获得了所有参与者的书面知情同意。)

可以看出类似这种标准语句的用法是非常普遍的, 如果在写作时参考模板却不做改写, 就会和已发表论文重复。

此外, References 参考文献列表中的参考文献格式很大概率会与前人重复, 除非你是第一次引用相同的文献。因此专业的查重软件在查重时就会排除参考文献进行查重。例如权威英文论文查重软件 iThenticate 就可以在查重时设置自动排除参考文献列表的选项。

以上就是论文写作中容易造成中英文论文无意重复的六大「罪魁祸首」, 在写作论文时, 我们需要有意识地注意它们。

参考文献

[1] 吴志根. 国际高水平 SCI 论文写作和发表指南[M]. 杭州: 浙江大学出版社, 2019: 128 - 134.

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