

XINCHECK® 文本查重

产品官网: https://www.xincheck.com

联系方式: 18701138792/18810760681



北京芯锋科技有限公司

Beijing XinFeng Technology Co., Ltd.

查重指标



- ✓ 查重报告编号
- ✓ <u>总字数</u>
- ✓ 重复字数
- ✓ 总文字复制比
- ✓ 总段落数
- ✓ 疑似段落数
- ✓ 疑似段落最大重合字数
- ✓ 疑似段落最小重合字数
- ✓ 单篇最大重复字数
- ✓ 单篇最大文字复制比
- ✓ 前部重合字数
- ✓ 后部重合字数



查重指标



SDK可以识别文本的前言目录部分和参考文献部分,前言目录部分将作为查重报告分块的第一块,参考文献部分将作为查重报告分块的最后一块。对于中间的正文部分,默认约每1万字划分为一块¹。

查重报告展示三大部分内容:

- 1、基本信息,包括文本标题、作者、查重报告编号、查重时间等。
- 2、整体检测指标,包括总文字复制比、单篇最大文字复制比、重复字数、总字数、单篇最大重复字数等等。
- 3、全文标明引文或原文对照结果。



¹SDK通过关键词进行前言目录、参考文献部分的识别,不能保证识别率。识别的成功与否不会影响总文字复制比的计算。

查重指标 查重报告编号



由字母、数字、符号组成的字符串。 SDK在单独使用过程中生成的查重报 告是唯一的,如果同时使用多个SDK, 查重报告编号无法保证唯一。

如图所示,查重报告编号为 2021D_16401615962930000000000 0000002449484。



查重指标 总字数



指被检测文本的总有效字符数2。

如图所示,被检测文本《地下室工程施工组织设计》的总字数为43383。



2查重报告所指的字数均为"有效字符数",在去除待查文件中多余的空行、空格、特殊符号后再查重。

查重指标 重复字数



指系统计算出的被检测文本与比对 库中文本比对后,出现重复的总字 符数。

如图所示,被检测文本《地下室工程施工组织设计》共计 39827 个字符数与比对库中的文本重复。



查重指标 总文字复制比



指被检测论文总的重复字符数占全 文字符数的比例,即重复字数除以 总字数减去参文献部分的字符数得 到的比值。通过该指标,可以直观 了解重复字数在该被查文本中所占 比例。

以被检测文本《地下室工程施工组 织设计》为例



查重指标 总段落数



如果被检测文本被进行了分块处理,则代表被检测文本被分成的块数。如果被检测文本没有被进行分块处理,该值为1。

以被检测文本《地下室工程施工组织设计》为例,系统将其划分为目录摘要及正文的五部分,共六块,即总段落数为 6。



查重指标 疑似段落数



指存在文字复制行为的文本块的数量。

以被检测文本《地下室工程施工组织设计》为例,系统将目录摘要即正文共划分为六个部分,每个部分均存在文字复制行为(即重复率/重复字符数不为 0),故疑似段落数为 6。



查重指标 疑似段落最大重合字数



所有存在重复的文本块中,重复字 数最多的段的重复字符数。

以被检测文本《地下室工程施工组织设计》为例,在系统划分的六个部分(均为存在复制行为的疑似段落)中,正文第 2 部分重复字符数最多,为 10242 个字符数。



查重指标 疑似段落最小重合字数



所有存在重复的文本块中,重复字 数最小的段的重复字符数。

以被检测文本《地下室工程施工组织设计》为例,在系统划分的六个部分(均为存在复制行为的疑似段落)中,目录摘要等重复字符数最少,为 1967 个字符数。



查重指标 单篇最大重复字数



指被检测文本与比对库中文本比对 后,得到的所有的相似文献中,重 复字数最多的那一个文献的重复字 数。

以被检测文本《地下室工程施工组织设计》为例,在比对库中所有与该文本相似的文献中,"最相似"的一篇文献与该文本有 39827 个字符数相同。



查重指标 单篇最大文字复制比



指被检测文本与比对库中文本比对后,得到的所有的相似文献中,重复字数最多的那一个文献的重复字数占待查文本总字数的比率。

以被检测文本《地下室工程施工组织设计》为例,单篇最大文字复制比为

单篇最大文字复制比(91.8%)

= <u>单篇最大重复字数 (39827)</u> 总字数 (43383)



查重指标 前部重合字数



指被检测文本的前言目录部分以及正文部分约前50%部分的重复字符数。

以被检测文本《地下室工程施工组织设计》为例,该文本的前部大约为目录摘要等和正文的第1部分、第2部分、第3部分、即

前部重合字数(28503)

- ≈ 目录摘要(1967) + 第1部分(8232)
- + 第2部分(10242) + 第3部分(8062)



查重指标 后部重合字数



指被检测文本的参考文献部分以及正文部分约后50%部分的重复字符数。

以被检测文本《地下室工程施工组织设计》为例,该文本的前部大约为正文的第 4 部分、第 5 部分,即

后部重合字数(11324)

≈ 第4部分(8498) + 第5部分(2826)



Thank you

