

第二章 Word高级搜索与替换

本章所有样本和例子文本请见 “\数据样本\第二章” 目录

一 Word搜索模式简介

1.1 普通搜索 Ctrl+F

仅能精确搜索，搜索结果与搜索内容严格匹配，缺乏灵活性。

1.2 高级搜索 Ctrl+H

1.2.1 两种模式

a. 普通高级搜索

- 能搜索格式（字体、字号、颜色等）、域（脚注、尾注标记等）
- 能搜索不可见字符（回车符、制表符、分节符、分页符等）
- 能进行简单的模式匹配搜索（如能搜索 任意字符、任意数字 等）

b. 通配符高级搜索

- “通配符”搜索能匹配模式，实现相对模糊的搜索功能；

二. Word高级检索之普通高级搜索

Word高级检索还支持通过文本格式搜索，这是Word优于其他文本编辑器的一个重要特点。

打开方法：










普通高级搜索，打开【替换】对话框的两种方法

- 【开始】选项卡中【替换】按钮；
- Ctrl+H


格式搜索：

按 **Ctrl+H** 进入替换对话框后，点左侧 **更多** 按钮，展开高级选项，在左下角的 **格式** 按钮中设置格式（支持字体、段落等等格式内容）后搜索即可。搜索内容框中不输入搜索内容，只设置格式，将会搜索到所有符合格式的内容，如包括内容，则搜索同时满足关键词和格式要求的结果。

普通高级搜索功能简介：

- 不可见字符搜索（段落标记：、制表符：、分页符：）
- 图片搜索：
- 脚注（）、尾注标记（）
- 简单的模式匹配匹配搜索（任意字符：，任意字母：，任意数字：）

练习

1. 在 **说文解字样本.doc** 中使用格式查找功能，选中所有字头。
2. 在 **说文解字样本.doc** 中通过格式查找功能选中所有  标记的内容。



三 Word高级检索之通配符搜索

3.1 打开方法：

- 打开搜索与替换对话框的两种方法
 - 【开始】选项卡中【替换】按钮；
 - **Ctrl+H**
- 通配符搜索，【更多】——勾选“使用通配符”

3.2 通配符搜索：

3.2.1. 搜索任意字符

- 通配符：
- 说明：一个代表一个任意字符，“任意字符”可以是字母、数字、标点、中文字符等可见字符以及制表符、回车符等不可见字符。

3.2.2 零个或多个字符

- 通配符：

- **说明**：当要搜索的模式中某一位置可能出现字符，也可能不出现的时候使用
- **例**：`[a-z]*\@163.com` 能匹配到 `a@163.com` 也可以匹配到 `abd@163.com`

3.2.3 范围内字符

- **通配符**：`[]`
- **两种模式**：
 - 给定范围，如 `[0-9]` 匹配所有数字，`[a-z]` 匹配所有小写字母，`[一-𠂇]` 匹配 CJK 基本平面所有汉字字符。
 - 列举所有指定字符：如 `[, . ? !]` 可匹配方框内4种标点
- 两种模式可以混合使用，如 `[a-z3]`，匹配26个小写字母或者数字3

3.2.4 非

- **通配符**：`[!]`
- **说明**：配合范围内字符 `[]` 使用，即非的是括号内的字符。可理解为：范围内字符前如有感叹号（英文）`!` 则搜索这些字符之外的内容。
- **例**：`[!0-9a-z,]` 搜索数字、小写字母，英文逗号之外的字符。

3.3.5 前面一个或多个

- **通配符**：`@`
- **说明**：常与范围内字符配合使用，用于指定范围内字符出现的次数
- **例**：`[a-z]@` 匹配任意长度的由小写字母组成的字符串。

3.3.6 前面内容的出现次数

- **通配符**：`{ }`
- **说明**：精确指定前面的字符（或范围内字符）出现的次数。
 - 可以指定确切的出现次数，如 `a{3}` 相当于搜索 `aaa`
 - 可以指定出现的次数范围，如 `从?{1,2}`，可以搜到 `从某`，也可以搜到 `从某某`
 - 可以只指定次数的上限或者下限，如 `{3,}` 至少出现3次，`{, 4}` 最多出现4次。

3.3.7 不可见字符的搜索

- **回车符**：`^13`
- **制表符**：`^t` 或者 `^9`

3.3.8. 转义符

- **通配符：** \
- **说明：** 由于高级检索模式中将 *?@ [] () { } 等字符作为模式匹配标记，在通配符模式下，这些字符有特殊的意义，直接搜索这些字符是搜索不到的。为搜到这些字符，需要使用所谓“转义符”，告诉计算机我们需要搜索的就是这个字符本身。
- **例：** \? 可搜索到问号

3.3 搜索范围和高亮显示结果

3.3.1 搜索范围（在以下项中查找）

- **主文档：** 默认搜索模式，会在当前打开文档的全文中进行搜索。
- **当前所选内容：** 如只想在文档中指定范围内进行搜索，可先选中该范围，然后选择在当前所选内容中搜索即可。当没有内容被选中时，此选项不会出现。

指定搜索范围点击搜索后，所有搜索结果都会被选中，此时按**Ctrl+C**，可以将所有结果复制出来。

3.3.2 突出显示

- **阅读突出显示：** 通过该选项可以将所有符合条件的结果高亮显示，方便对符合条件的搜索结果进行全局观察（普通查找模式下，默认即是此效果）。

3.4 练习

1. 在 说文解字样本.doc 文本中找出所有 某某切 的反切形式，并将结果复制到新文档。
2. 在 说文解字样本.doc 文本中找出所有反切上字为 古 的反切。
3. 在 说文解字样本.doc 文本中找出用 《 》 标记的引书。
4. 在 说文解字样本.doc 文本中选中所有方括号 [] 包裹的内容。
5. 将 说文解字样本.doc 文本中的所有字头找出并复制到新文档。
6. 从 人民日报标注语料库样本.doc 文件中搜索选中所有的时间、文件戳，如 19980101-01-001-001。
7. \唐诗三百首\001.txt 出现了哪些诗人？请使用Word查找替换及Excel的筛选功能解答。

四. Word替换

通过Word替换功能，可以对数据进行一些转换处理，以下通过具体例子进行展示。

4.1 删除内容

搜索指定内容，替换框中不填，替换后，即可删除搜索到的内容。

例：搜索空格，替换为空，可删除所有空格。

题：如何删除所有多余的空行？

4.2 要查找的表达式

当要将查找内容中的一部分内容进行增、删、替换等操作时，需要对查找内容的部分进行定位，此时可使用“要查找的表达式”功能进行处理。

方法：

1. 在查找框中，使用 `()` 标记（表达式标记）需处理的部分，可根据需要重复使用表达式标记。
2. 在替换框中，使用 `\n` 标记（要查找的表达式，n为表达式序号），指代查找框中的表达式。
3. 如要删除某一表达式内容（如第3段表达式），在替换框中不出现该指代符即可（如不出现 `\3`），如要替换，同理。
4. 表达式可重复引用。例见 `搜索表达式重复引用例.doc`。

例子：

1. 在 `人民日报标注语料库样本.doc` 文件中将时间、文件戳（形如 `19980101-01-001-001`）转换为如下形式：`19980101:01:001:001`。
2. 在 `人民日报标注语料库样本.doc` 文件中，删除时间、文件戳中的第3段，即变为如下形式：`19980101:01:001`。

五. 综合练习

5.1 唐诗三百首的规整化

要求：

1. 删除 `返回首页` 等无关信息；
2. 将诗文内容变为一行；
3. 删除诗文标题和诗文内容之间的空行。

最后效果示例：

001张九龄：感遇四首之一

孤鸿海上来，池潢不敢顾。侧见双翠鸟，巢在三珠树。矫矫珍木巅，得无金丸惧。美服患人指，高明逼神恶。今我游冥冥，弋者何所慕。

5.2 一切經音義引書引人索引.doc 的格式化

要求：

1. 引文出处格式修改为 四号 宋体 ；
2. 引文标题格式修改为 三号 黑体 ，大纲级别 2级 ；
3. 引文出处为 无 的，用红色标记。