**FFReader-GB18030-2022测试使用指南**

刘德位

793554262@qq.com

# 基础要求

更新FFReader主程序版本最低1.9.25版本，请前往<https://www.ffreader.cn/download> 下载最新版本的FFReader

操作系统最低要求

WindowsXP ，请下载32位版本的FFReader

Windows7以及以上版本的操作系统，请下载64位版本的FFReader（不建议在Win7以及以上版本的操作系统使用32位FFReader，这将导致工具很难支持大文件）

# 字体补充

截止到2023年7月，目前大部分操作系统还没有很好的支持GB18030-2022所有字符的显示，主要原因是系统内的字库信息不全，为了快乐的解决此问题，请前往<https://www.ffreader.cn/download> 下载生僻字补充包，包内<字体-相关补充字体>目录下包含相关补充字体文件，请参考目录内的Readme文件安装后重新打开FFReader，即可支持到GB18030-2022所有字符（不包含第二次意见征求稿的补充汉字）

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

# 验证FFReader在生僻字解析能力

按上一章节补充安装字体后，为了验证FFReader在你的电脑上，是否可以正常解析并显示GB18030-2022的所有汉字，建议使用生僻字补充包中的测试文件验证你的FFReader使用可以解析并展示相关生僻字

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

如图，这两个文件是生僻字清单，这里分别有UTF-8和GB18030两种编码方式的文件，内容是一样的

请注意：如果你电脑上的文本编辑器对GB18030-2022支持能力有限，那么请不要轻易编辑GB18030版本的测试文件，这可能导致文件被破坏乱码(一般情况下，编辑器打开各种编码的文件会解码为该程序内部对应编码进行展示处理，保存时重新编码为GB18030，因为GB18030编码需要映射表实现，老版本的文本编辑器可能因为缺少GB18030-2022的支持导致再次保存时出现乱码情况)

如下是验证情况

图形用户界面, 文本, 应用程序

描述已自动生成

如截图所示，在安装完毕相关字体后，FFReader程序可以解析到CJK-F汉字，CJK-F后面的汉字，因为缺少字体支持，显示为方块，后续相关字体文件更新后，即可显示正常，不要求再次更新FFReader

# 相关编码和GB18030-2022支持的关系

国标要求，2023年8月1号就要支持GB18030-2022，是说大家要把系统改造成GB18030编码么？

国标要求，相关信息系统支持GB18030-2022编码的相关汉字，并不是说一定要求信息系统数据库、程序全部使用GB18030编码，这里是等效支持的概念。目前大部分程序和字体，都是使用UNICODE作为内码处理字符的，UNICODE码作为一个标准字符编码方式一般并不作用于数据库、文件。我们一般使用UTF-8,GB2312,GBK,GB18030这种编码方式存储数据，文件，一个字符在UNICODE,UTF-8,GBK,GB1830等编码方式下具有唯一（不考虑PUA双码字）对应关系，比如下面截图的“我”字

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

那么，在各个编码等效的情况下，如果你的程序使用UTF-8编码存储数据，只要支持的字符范围涵盖到GB18030-2022即可，UNICODE 15.0的范围约等于GB18030-2022（因为UTF-8是对UNICODE编码的一种变长表示，UNICODE和UTF-8不需要复杂的程序支持（通用公式转换），所以我们一般说UNICODE版本升级对UTF-8是无感的，只要字库支持响应版本UNICODE新增汉字，UTF-8即自动支持）。

千千秀字，汉字编码查询网站<https://www.qqxiuzi.cn/bianma/zifuji.php>

既然说各种字符编码是一一对应的，那我为什么要升级GB18030-2022？

1：各种字符编码的老版本相较于新版本，都是缺少大量新增字符的，比如目前大家中文系统广泛使用的GBK编码，只支持几千个汉字，很多生僻字都是不支持的

2：强制大家升级程序字符支持能力，扩充生僻字处理能力，即使你程序内使用的是UTF-8，但是各个信息系统交换文件数据时，协议要求使用GB18030编码（比如开放式基金交换协议），这里可能引入一个大转小的问题，假如你的数据库编码是UTF-8，支持生僻字，但是导出GB18030编码的文件时，因为GB18030版本较老，该版本GB18030无该字对应的编码，导致UTF-8转GB18030时，转换成了乱码（通常字符在两个编码之间转换时，大编码范围转小编码范围，会将对应字符转换成问号或者多个乱码字符）

3：取消之前不明确的PUA双码字编码，之前因为各种历史原因，有不少汉字被重复编码，这导致大量程序出现处理异常，比如下面这个字



同字形的情况下UNICODE码却有两个，这种情况下比如身份证系统使用E863编码改字，银行使用4DAE，则可能遇到银行获取身份证信息进行实名验证时，比对失败（字形看起来一样，但是字符编码值不同，导致程序认为不是同一个字），导致无法开户，GB18030-2022强制要求使用非PUA编码，试图解决这种问题（因为大量的历史原因，这种情况可能不会这么快的解决！，因为各个系统内存在大量的老编码的信息），常用的PUA双码字，可参考<常见的60个有PUA的双码汉字.txt>文件

4：强制大家升级数据库字符集，当前大量的信息系统采用mysql utf8mb3字符集，该字符集使用3字节编码，导致4字节编码汉字无法存储，大量生僻字在UTF-8编码下，是4字节编码的，所以使用utf8mb3的表需要升级为utf8mb4

# 关于GB18030-2022测试的建议

1：如果你的信息系统计划对PUA强制转码成正式码，则可使用FFReader的相关功能识别PUA信息，比如

图形用户界面, 应用程序

描述已自动生成

在相关字段上选择红框标记的功能

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

这里会列出该汉字的信息，并提示该字为PUA码，建议使用正式编码（受限于本工具收集的PUA双码汉字信息，可能该提示不全）

2：如果你的信息系统有文件交换功能，可以验证导出的文件是否可以在FFReader中正确的显示，为了测试应用系统对生僻字的解析能力，建议在CJK扩展A-F中各抽选一些汉字测试，确保你的应用系统都可以正常导出包含该文字的文件，并且合作方发送的包含该文字的文件可以正常读取，下面是一些示例

图形用户界面, 应用程序, Word

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

# 生僻字很复杂，看不清怎么办

你可以在FFReader中选择该字段，右键使用放大镜查看，该功能将该字段信息，使用60px超大字体显示

图形用户界面, 应用程序

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

使用该功能可以清晰的看到该字的结构信息