



Unicode
GB 18030

-
- Unicode
- GB 18030
- GBK Big5 UTF-8
-
-
-
- PUA
-
-

“ ”

1.

Unicode / GB 18030
2.
3.

		3
?		2
	GBK UTF-8	2
		1 2
		3

- Unicode

Unicode



“

B



Unicode

U+

U+4E2D

Unicode

17

65,536

- 0
- 2

3
- 15

16

PUA



Unicode 17.0

2025

	4E00–9FFF	20,992	0
A	3400–4DBF	6,592	0
B	20000–2A6DF	42,720	2
C	2A700–2B73F	4,160	2
D	2B740–2B81F	222	2
E	2B820–2CEAF	5,774	2
F	2CEB0–2EBEF	7,473	2
I	2EBF0–2EE5F	622	2

[illegible]

GB 18030

--	--	--	--

GB 2312 □ GBK□

- 2000  2005 
 -  **GB 18030-2022** 2022 7  2023 8 1 
 -  87,887  1.7
- 

--	--	--	--	--	--

 3 9

項目	内容	備考
① 1	XXXXXXXXXX XXXXXXXXXX CJK A	XXXXXX
② 2	① 1 XXXXXXXXXXXXXXXXX XXXXXX 1	XXXXXXXXXXXXXXXX XXXXXX
③ 3	② 2 XXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXX

--	--	--	--	--	--	--

1

2024 9 29 1 Unicode G H I 3
9793 1 2



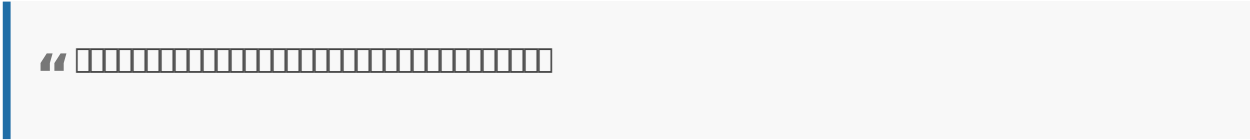
- 3 Unicode
- 2022
- 4
- PUA





GBK与 Big5

UTF-8



- 每个 字节 代表一个 字符
- 每个 字节 代表一个 字符

Unicode与 GB 18030
UTF-8与 UTF-16与 GBK

每个 字节 UTF-8 3 字节 GBK 2 字节



名称	字符数	备注
GB 2312	6763 个	早期标准
GBK	21000 个	GB 2312 的扩展 Windows 默认编码
GB 18030	21000 个 Unicode 兼容	GBK 的扩展 GB 18030 国家标准
Big5	13000 个	繁体中文标准
UTF-8	21000 个 Unicode 兼容	国际通用标准



- 每个 字节 代表一个 字符 与 GBK 兼容
- 每个 字节 代表一个 字符 UTF-8 与 GBK 兼容

• CSV Excel UTF-8 Excel UTF-8 BOM UTF-8



1. UTF-8 UTF-8
2. `<meta charset="utf-8">` `<head>`
3. UTF-8 MySQL utf8mb4
4. GB 18030 GBK GBK



UTF-8 UTF-8 UTF-8 UTF-8 UTF-8 UTF-8 UTF-8 UTF-8

UTF-8

--	--	--	--	--	--	--

“ []
[]

[illegible]

U

- 

- 

- 

[illegible]

- **Windows**   Unicode
- **macOS**   Control + Command + 

[illegible]

Unicode Unicode Unicode

- **Microsoft Word**                    

- **macOS** 10.15 → 10.14 → 10.13 Unicode 10.13 Option
- 10.13 / 10.14 10.15 HTML 10.15 CSS 10.15 JavaScript 10.15 "u{20B9F}" 10.15



- 10.13 10.14 10.15
- 10.13 10.14 10.15

10.13 PUA 10.13

10.13 10.14



10.13 Unicode 10.13

- 10.13 10.14 10.15
- 10.13 10.14 10.15

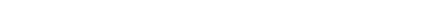
10.13 10.14

10.13

IVS 

U+E0100 □ U+E01EF



-  Unicode  IVD
-  Adobe-Japan1 Hanyo-Denshi Moji_Joho MSARG KRName CAAPH
 Moji_Joho  2026  8 
- 




IVS



- 1.  
- 2.  
- 3.  



IDS



-  
-  



Unicode 15.1





“

App



Windows macOS



```
body {
  font-family:
    " ", /*  */
    "Plangothic P1", /*  */
    "Plangothic P2",
    sans-serif;
}
```



-  
-   P1 P2

-  **unicode-range**   **@font-face**


```
@font-face {
  font-family: "FallingBalls";
  src: url("/fonts/fallback-extb.woff2") format("woff2");
  unicode-range: U+20000-2A6DF, U+2A700-2EE5F, U+30000-3347F;
}
```

-  
-    



- Word WPS 
- PDF 
- 
 







1.   Unihan 
2.  
3. 

[illegible]

1



OpenType





项目	说明	备注
数据库	MySQL 支持 utf8 和 utf8mb3 编码，默认使用 utf8mb4 编码，长度为 4 字节。	数据库默认使用 utf8mb4 编码。
字符串	使用 Unicode 编码，范围 [\u4e00-\u9fa5]，长度为 4 字节。	JavaScript 使用 Unicode 编码，范围 /\p{Script=Han}/u。
文本	使用 UTF-16 编码，长度为 2 字节。	文本默认使用 UTF-16 编码。
图片	使用 UTF-8 编码，长度为 1 字节。	图片默认使用 UTF-8 编码。
文件	使用 GBK 编码，长度为 2 字节。	文件默认使用 UTF-8 编码，范围 18030 GB。
链接	使用 UTF-8 编码，长度为 1 字节。	链接默认使用 UTF-8 编码，范围 18030 GB。
PDF	使用 PDF 编码，长度为 1 字节。	PDF 默认使用 PDF 编码，长度为 1 字节。



JavaScript 使用 UTF-8 编码，长度为 1 字节。

```
const isHan = (s) => /^\\p{Script=Han}+$/u.test(s);
isHan(" "); // true
U+20BB7 B
```

```
[...""].length; // 2
"".length 3
```

MySQL

```
CREATE TABLE person (
  name VARCHAR(64) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci
);
```

utf8mb4

- A
- B 4
- Excel

GB 18030-2022 3 [GB 18030](#)

“

한글	한글
Unicode	
Unicode	
Unicode UniHan	Unicode

1. Unicode U+20BB7
2.
3.
4.
5.

Issues

--	--	--	--	--

[illegible]

license@wenfeng.org

[illegible]

--	--	--	--	--	--

Unicode

IRG

[illegible]

-

-

□□□□ PUA□□□□□□□□ □