

1 char 数据类型概述

2 Character 数据类型概述

3 Character 方法简述

3.1 是否有效 或 为 Surrogate 类型

3.2 特定下标(之前)的 code point 值

3.3 codepoint 与 surrogate 的转换

3.4 因为 surrogate 导致的统计个数

3.5 codepoint 代表的类型 General category

3.6 杂

3.7 toLowerCase / toUpperCase / toTitleCase

参考

## 1 char 数据类型概述

---

- 码点 ( code point ) : unicode 字符集中字符对应的数值
- 在Java中,char 数据类型 ( 和 Character 对象封装的值 ) 基于原始的 Unicode 规范 ( 字符集 )
- char表示 UTF-16 编码的代码单元 (code unit)
- 对于0号平面(BMP),一个码点(code point)对应 1 个 code unit —— 1 个 char 对应一个 unicode 字符集中的字符
- 对于辅助平面,一个 code point 将会对应 2 个 code unit—— 2 个 char 对应一个 unicode 字符集中的字符

## 2 Character 数据类型概述

---

- Character在 jdk8中, 基于版本Unicode6.0.2 标准
- Character 类的方法和数据是通过 UnicodeData 文件中的信息定义的
- 此文件指定了各种属性, 其中包括每个已定义 Unicode 代码点或字符范围的名称和常规类别(General category)
- 详情可查阅 <http://www.unicode.org>

## 3 Character 方法简述

---

### 3.1 是否有效 或 为 Surrogate 类型

返回类型	方法名	简述	补充
static boolean	isValidCodePoint(int var0)	是否为有效的 codePoint	(var0 >>> 16) < 17
static boolean	isBmpCodePoint(int var0)	是否在第 0 平面	var0 >>> 16 == 0
static boolean	isSupplementaryCodePoint(int var0)	是否在辅助平面	var0 >= 0x10000 && var0 < 0x110000;
static boolean	isHighSurrogate(char var0)	是否是高代理	var0 >= '\ud800' && var0 < '\udc00'
static boolean	isLowSurrogate(char var0)	是否是低代理	var0 >= '\udc00' && var0 < '\ue000'
static boolean	isSurrogate(char var0)	是否是代理	综上
static boolean	isSurrogatePair(char var0, char var1)	是否是代理对	isHighSurrogate(var0) && isLowSurrogate(var1)

## 3.2 特定下标(之前)的 code point 值

返回类型	返回类型方法名	简述
static int	codePointAt(CharSequence seq, int index)	获取指定位置的 codePoint
static int	codePointAt(char[] var0, int var1)	同上
static int	codePointAt(char[] var0, int var1, int var2)	同上，加范围限制
static int	codePointBefore(CharSequence var0, int var1)	获取前一个位置的 codePoint
static int	codePointBefore(char[] var0, int var1)	
static int	codePointBefore(char[] var0, int var1, int var2)	

## 3.3 codepoint 与 surrogate 的转换

返回类型	方法名	简述
static char[]	toChars(int var0)	将 codePoint 转为数组
static int	toChars(int var0, char[] var1, int var2)	保存到数组的指定位置
static int	toCodePoint(char var0, char var1)	将 surrogatePair 合成为 codePoint,这之前要检查 surrogatePair 类型
static char	highSurrogate(int var0)	返回代码点的高代理
static char	lowSurrogate(int var0)	返回代码点的低代理

### 3.4 因为 surrogate 导致的统计个数

返回 类型	方法名	简述	补充
static int	charCount(int var0)	codePoint 对应的 char 的个数	
static int	codePointCount(CharSequence var0, int var1, int var2)	v1-v2 之间的 <b>codePoint</b> 的个数	不是 char 的个数
static int	codePointCount(char[] var0, int var1, int var2)	同上	
static int	offsetByCodePoints(CharSequence var0, int var1, int var2)	返回偏移指定个数的 <b>codepoint</b> 后的索引	由于 codepoint 可能对应 2 个 char，索引不能简单的直加
static int	offsetByCodePoints(char[] var0, int var1, int var2, int var3, int var4)	同上，加限制了	

### 3.5 codepoint 代表的类型 General category

返回类型	方法名	简述	补充
static boolean	isLowerCase(int var0)	小写?	
static boolean	isUpperCase(int var0)	大写?	
static boolean	isTitleCase(int var0)	首字母大写?	
static boolean	isDigit(int var0)	数字?	
static boolean	isDefined(int var0)	被定义为 Unicode 中的字符?	
static boolean	isLetter(int var0)	字母?	
static boolean	isLetterOrDigit(int var0)	字母或数字?	
static boolean	isJavaLetter(char var0)		isJavaIdentifierStart(var0)
static boolean	isJavaLetterOrDigit(char var0)		isJavaIdentifierPart(var0)
static boolean	isAlphabetic(int var0)	比 isLetter 大一些	确定指定的字符 (Unicode代码点) 是否是字母表
static boolean	isIdeographic(int var0)		是否是Unicode标准定义的CJKV (中文, 日文, 韩文和越南文) 表意文字
static boolean	<b>isJavaIdentifierStart</b> (int var0)	The "Java letters" include uppercase and lowercase ASCII Latin letters A-Z (\u0041- \u005a), and a-z (\u0061-\u007a),	and, for historical reasons, the ASCII underscore (_, or \u005f) and dollar sign (\$, or \u0024).
static boolean	isJavaIdentifierPart(int var0)	比上面多了数字	<a href="#">参考 3.8 章节</a>
static boolean	isUnicodeIdentifierStart(int var0)	是否允许作为 Unicode 标识符中的首字符?	isLetter    LETTER_NUMBER
static boolean	isUnicodeIdentifierPart(int var0)	是否允许作为 Unicode 标识符中的首字符以外的字符?	不懂

返回类型	方法名	简述	补充
static boolean	isIdentifierIgnorable(int var0)	是否是 Java 标识符或 Unicode 标识符中可忽略的一个字符?	不懂
static boolean	isSpaceChar(int var0)	unicode 中的空白字符?	
static boolean	isWhitespace(int var0)	Java 中的空白字符?	
static boolean	isISOControl(int var0)	是否是ISO控制字符	
static boolean	isMirrored(int var0)	是否镜像	

### 3.6 杂

返回类型	方法名	返回类型	简述	补充
static String	String getName(int codePoint)	static String	返回codepoint 所属类型的名称	
static int	getNumericValue(int codePoint)	static int	返回指定的 Unicode字符代表的罗马数值	字符'\u216C'(罗马数字50) 将返回一个int 值50
static int	digit(int codePoint, int radix)	static int	codePoint 对应 radix 进制的数值	因为 0-9 + a-z , 总数是 36 , 所以 $2 \leq \text{radix} \leq 35$ ; digit('h',20)==17
static int	char forDigit(int digit,int radix)	static int	digit 对应的进制的字符	forDigit(17,20)=='h'
static Character	valueOf(char c)	static Character	返回包装类型	
static char	reverseBytes(char ch)	static char	返回通过反转指定的 char值中的字节顺序获得的值	
static int	getType(int var0)	static int	返回 codepoint 类型	
static byte	getDirectionality(int var0)	static byte	给定字符的 Unicode方向性属性	

### 3.7 toLowerCase / toUpperCase /toTitleCase

返回类型	方法名	简述
static int	toLowerCase(int codePoint)	返回对应的小写的 codepoint
static int	toUpperCase(int codePoint)	返回对应的大写的 codepoint
static int	toUpperCaseEx(int codePoint)	如果错误，返回 ERROR 代码
static char[]	toUpperCaseCharArray(int codePoint)	返回 char[] 形式
static char	toTitleCase(int codePoint)	不懂

## 参考

---

jdk 1.8\_171

[官方 API](#)

[基础数据类型之Character](#)