

字符串基础

定义

字符集

一个**字符集** Σ 是一个建立了**全序**关系的集合，也就是说， Σ 中的任意两个不同的元素 α 和 β 都可以比较大小，要么 $\alpha < \beta$ ，要么 $\beta < \alpha$ 。字符集 Σ 中的元素称为字符。

字符串

一个**字符串** S 是将 n 个字符顺次排列形成的序列， n 称为 S 的长度，表示为 $|S|$ 。

如果字符串下标从 1 开始计算， S 的第 i 个字符表示为 $S[i]$ ；

如果字符串下标从 0 开始计算， S 的第 i 个字符表示为 $S[i - 1]$ 。

子串

字符串 S 的**子串** $S[i..j]$ ， $i \leq j$ ，表示 S 串中从 i 到 j 这一段，也就是顺次排列 $S[i], S[i + 1], \dots, S[j]$ 形成的字符串。

有时也会用 $S[i..j]$ ， $i > j$ 来表示空串。

子序列

字符串 S 的**子序列** 是从 S 中将若干元素提取出来并不改变相对位置形成的序列，即 $S[p_1], S[p_2], \dots, S[p_k]$ ， $1 \leq p_1 < p_2 < \dots < p_k \leq |S|$ 。

后缀

后缀 是指从某个位置 i 开始到整个串末尾结束的一个特殊子串。字符串 S 的从 i 开头的后缀表示为 $Suffix(S, i)$ ，也就是 $Suffix(S, i) = S[i..|S| - 1]$ 。

真后缀 指除了 S 本身的 S 的后缀。

举例来说，字符串 `abcbcd` 的所有后缀为 `{d, cd, bcd, abcd, cabcd, bcabcd, abcbcd}`，而它的真后缀为 `{d, cd, bcd, abcd, cabcd, bcabcd}`。

前缀

前缀 是指从串首开始到某个位置 i 结束的一个特殊子串。字符串 S 的以 i 结尾的前缀表示为 $Prefix(S,i)$ ，也就是 $Prefix(S,i) = S[0..i]$ 。

真前缀 指除了 S 本身的 S 的前缀。

举例来说，字符串 `abcbcd` 的所有前缀为 `{a, ab, abc, abca, abcab, abcab, abcabcd}`，而它的真前缀为 `{a, ab, abc, abca, abcab, abcab}`。

字典序

以第 i 个字符作为第 i 关键字进行大小比较，空字符小于字符集内任何字符（即： $a < aa$ ）。

回文串

回文串 是正着写和倒着写相同的字符串，即满足 $\forall 1 \leq i \leq |s|, s[i] = s[|s| + 1 - i]$ 的 s 。

汉明距离

汉明距离 是两个等长字符串之间的距离，它表示两个长度相同的字符串对应位字符不同的数量。

我们可以简单的认为对两个串进行异或运算，结果为 1 的数量就是两个串的汉明距离。

字符串的存储

- 使用 `char` 数组存储，用空字符 `\0` 表示字符串的结尾（C 风格字符串）。
- 使用 C++ 标准库提供的 `string` 类。
- 字符串常量可以用字符串字面量（用双引号括起来的字符串）表示。

参考资料与注释

- [后缀数组](#) by. 徐智磊

🔧 本页面最近更新：2025/4/9 17:36:28，[更新历史](#)

✎ 发现错误？想一起完善？ [在 GitHub 上编辑此页！](#)

👤 本页面贡献者：Ir1d, minghu6, NachtgeistW, StudyingFather, countercurrent-time, Enter-tainer, H-J-Granger, ouuan, CCXXI, mgt, AngelKitty, cjsoft, diauweb, Early0v0, ezoixx130, GekkaSaori, Konano, LovelyBuggies, Makkiy, P-Y-Y, PotassiumWings, qinggniq, SamZhangQingChuan, sshwy, Suyun514, weiyong1024, current2020, GavinZhengOI, Gesrua, ghj1222, Great-designer, Haohu Shen, i-Yirannn, i-yyi, kxccc, lychees, Peanut-Tang, SukkaW,

Tiphereth-A, Xeonacid

© 本页面的全部内容在 [CC BY-SA 4.0](#) 和 [SATA](#) 协议之条款下提供，附加条款亦可能应用