

## Q01: 基础问题 1

**提示 (Prompt) :**

暴力算法与 Knuth-Morris-Pratt (KMP) 算法在子串搜索中的主要区别是什么？

**模型名称 (model\_name) :** ChatGPT

**模型版本 (model\_version) :** GPT-4

**回答 (response) :**

暴力算法和 KMP 算法的主要区别在于处理失配字符的方式。暴力算法从文本的每个位置开始检查模式匹配，如果出现失配则回退，导致最坏情况下的时间复杂度为  $O(mn)O(mn)O(mn)$ 。KMP 算法通过使用前缀函数 (Prefix Function)，避免不必要的比较，直接跳过已匹配的字符，从而实现线性时间复杂度  $O(n+m)O(n+m)O(n+m)$ 。

---

## Q02: 基础问题 2

**提示 (Prompt) :**

解释 KMP 算法中的前缀函数的作用。

**模型名称 (model\_name) :** ChatGPT

**模型版本 (model\_version) :** GPT-4

**回答 (response) :**

KMP 算法中的前缀函数用于计算模式串中每个位置的最长前缀长度，该前缀同时也是该位置之前子串的后缀。前缀函数帮助算法在发生失配时跳过无效的匹配，从而避免重复比较，加速字符串匹配过程。

---

## Q03: 细节问题 1

**提示 (Prompt) :**

与有限状态自动机 (FSA) 方法相比，后缀树算法如何改进子串搜索？

**模型名称 (model\_name) :** ChatGPT

**模型版本 (model\_version) :** GPT-4

**回答 (response) :**

后缀树算法构建了一个紧凑的字典树结构，能够在  $O(n)O(n)O(n)$  时间内表示字符串的所有后缀，支持高效的子串搜索、模式匹配和最长重复子串查找。相比之下，有限状态自动机 (FSA) 方法需要预先计算所有可能的状态转换，其时间复杂度为  $O(m^3|\Sigma|)O(m^3|\Sigma|)O(m^3|\Sigma|)$ ，占用大量内存。后缀树在空间利用上更高效，特别适用于处理包含重复子串的大型文本。

---

## Q04: 细节问题 2

**提示 (Prompt) :**

在动态规划算法（如 Needleman-Wunsch 和 Smith-Waterman）中，间隙惩罚的作用是什么？

**模型名称 (model\_name) :** ChatGPT

**模型版本 (model\_version) :** GPT-4

**回答 (response) :**

间隙惩罚用于在序列比对中惩罚插入和删除操作，防止算法过多地引入间隙 (Gap)。在 Needleman-Wunsch（全局比对）和 Smith-Waterman（局部比对）算法中，间隙惩罚平衡了匹配字符与插入/删除操作之间的权重。常见的间隙惩罚策略包括固定间隙惩罚和仿射间隙惩罚，后者在打开间隙时成本较

高，但延长间隙的成本较低。

---

## Q05: 前沿问题

### 提示 (Prompt) :

现代机器学习方法如何改进多序列比对 (MSA)，超越传统的动态规划方法？

**模型名称 (model\_name) :** ChatGPT

**模型版本 (model\_version) :** GPT-4

### 回答 (response) :

现代机器学习方法通过深度学习模型（如 Transformer 和循环神经网络 RNN）改进多序列比对 (MSA)。这些模型能够捕获序列中的远程依赖关系和上下文信息。例如，预训练模型如 AlphaFold 和 ESM 已在蛋白质结构预测领域取得了最先进的成果，说明类似模型可用于优化 MSA 任务。机器学习方法在处理复杂的序列比对时，可以减小计算复杂度，提供比传统动态规划算法更高的比对精度。