Software Engineering and Applications 软件工程与应用, 2023, 12(3), 578-585
Published Online June 2023 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2023.123056

Hans汉斯

数据结构中串的模式匹配算法演示系统的研究

刘城霞，宋泽昊

北京信息科技大学计算机学院，北京

收稿日期：2023年4月18日；录用日期：2023年6月23日；发布日期：2023年6月30日

摘　要

本文研究了数据结构中的朴素模式匹配算法(BF)和快速模式匹配算法(KMP)，并将其算法程序、中间变量、结果展示结合到一起，使用JavaSwing等进行相关图形界面开发，使得在同一界面上，不仅能展示BF算法和KMP算法的代码，并且能够显示字符串模式匹配算法执行过程中每一步的操作流程，还可以随时进行暂停、回退、继续等。该演示系统不仅可以帮助学生快速学习理解字符串模式匹配的原理，还可以有效地提高学生的实践能力，所见即所得地快速上手编程。

关键词

朴素模式匹配算法，快速模式匹配算法，JavaSwing

Research of a Demonstration System for Pattern Matching Algorithms of Strings in Data Structure

Chengxia Liu, Zehao Song

Computer School, Beijing Information Science and Technology University, Beijing

Received: Apr 18th, 2023; accepted: Jun. 23rd, 2023; published: Jun. 30th, 2023

Abstract

This article has studied the naive pattern matching algorithm (BF) and the fast pattern matching algorithm (KMP) in data structures, and combined their algorithm programs, intermediate variables, and result display. Using JavaSwing and other related graphical interface development, it is possible to not only display the code of the BF algorithm and KMP algorithm on the same interface, but also display the operation flow of each step in the execution process of the string pattern matching algorithm. It also can be pause, rollback, resume, and so on at any time. The demonstra-

文章引用: 刘城霞, 宋泽昊. 数据结构中串的模式匹配算法演示系统的研究[J]. 软件工程与应用, 2023, 12(3): 578-585.
DOI: 10.12677/sea.2023.123056

tion system can not only help students quickly learn and understand the principles of string pattern matching, but also effectively improve students' practical abilities, namely, WYSIWYG, and fast hands-on programming.

Keywords

Naive Pattern Matching Algorithm, Fast Pattern Matching Algorithm, JavaSwing

1. 引言

字符串(简称串)是一种重要的数据结构。随着计算机科学的发展，信息变得越来越多，而字符串构成的文本资料是信息的重要载体之一。想要在海量的文本资料中查找到所需信息，那就需要进行信息检索，而好的字符串模式匹配算法可以极大地提高检索的效率和质量。除此之外串的模式匹配算法还广泛应用于搜索引擎[1]、模式识别[2]、数据挖掘[3]、信息安全[4]、网络入侵检测[5]、数据压缩[6]、拼写检查[7]等领域中。

为更好地学习这些串的模式匹配算法，一种较好的方式就是通过动画演示来帮助理解，快速入门。演示系统可以图文并茂的展示算法的原理、过程及结果，既有助于教师生动的教学，又帮助学生直观地理解。目前也有许多相关的演示系统[8] [9] [10]，本文在研究了这些系统的基础上，结合学生的需要研究并开发了更全面、更简洁、更方便的串模式匹配算法演示系统。

2. 串的模式匹配算法

串模式匹配算法是指在主串中找到与模式串相同的子串，并返回其在主串中的所在位置。假设主串长度为 n，模式串长度为 m。在数据结构中，主要的模式匹配算法有 BF (Brute Force)算法和 KMP (由 D. E. Knuth，J. H. Morris 和 V. R. Pratt 提出)算法[11]。其中 BF 算法是简单的模式匹配算法，而 KMP 算法是一种改进的快速模式匹配算法。

2.1. BF 算法

BF 算法就是将主串中所有长度为 m 的子串依次与模式串对比，直到找到一个完全匹配的子串，或所有的子串都不匹配为止。BF 算法匹配原理如图 1 所示。

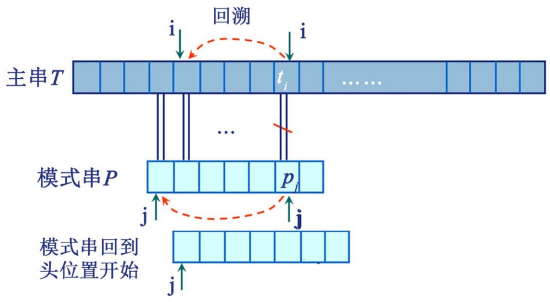

Figure 1. Matching principle of BF algorithm
图 1. BF 算法匹配原理

2.2. KMP 算法

KMP 算法，设 i 为主串中指向当前需要匹配字符的指针，j 为模式串中指向当前需要匹配字符的指针，当 i 和 j 所指向的字符不匹配时，主串指针 i 不进行回溯，只令模式串指针 j 回溯到模式串的最大公共前后缀+1 的位置，然后主串和模式串再进行模式匹配。具体过程如下：

1) 第 1 个元素匹配失败，匹配下一个相邻子串，j = 0，i++，j++；

2) 其他元素匹配失败，主串指针 i 不回溯，模式串指针 j 按 next 数组回溯，k = next[j]。其中 next[j]里面存放的是模式串位置 0 至位置 j − 1 的子串中最大公共前后缀长度。

KMP 算法匹配原理如图 2 所示。

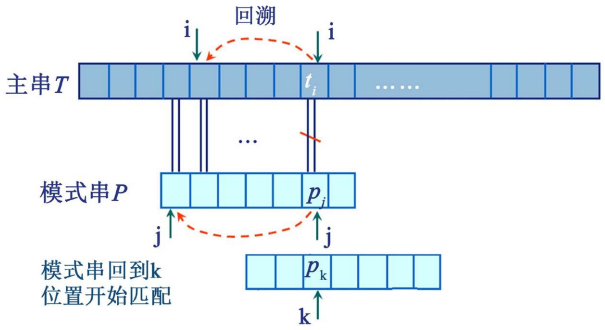

Figure 2. Matching principle of KMP algorithm
图 2. KMP 算法匹配原理

在 KMP 算法中，需要首先计算模式串中每一个位置 j 的字符和主串中 i 位置字符不匹配时，在主串不回溯(i 不变)的情况下主串 i 位置要和模式串的 k = next[j]位置进行新的匹配。而求 k = next[j]是与主串无关的，经过总结，可以得到 next[j]的计算公式为

$$\text{next}[j]=\begin{cases}-1 & j=0\\ k & 0\le k<j\text{且使用}"p_0\cdots p_{k-1}"="p_{j-k}\cdots p_{j-1}"\text{的最大整数}\end{cases} \quad \text{公式(1)}$$

3. 串的模式匹配算法演示系统设计

在演示程序中，程序的展示界面要能够显示 BF 算法和 KMP 算法的 Java 代码，并且要能够显示字符串模式匹配算法执行过程中每一步的操作流程，包括确定当前正在进行模式匹配的主串的子串，确定当前主串指针 i 和模式串指针 j 的所在位置，以及突出显示当前操作对应模式匹配算法的某一句代码等细节，让使用者能够清楚明白程序运行的实时情况。而且在演示程序中要提供简单方便的操作，如连续执行，单步执行，暂停等，有助于使用者学习理解字符串模式匹配算法。当演示程序正在演示字符串模式匹配算法的执行过程时，主要能够展示通过主串指针 i 和模式串指针 j 所指位置，对主串中的子串和模式串进行所指当前位置的模式匹配，如果所指当前位置的模式匹配成功，则指针指向下一个相邻位置再进行模式匹配；如果所指当前位置模式匹配失败，则主串指针 i 和模式串指针 j 回溯到由字符串模式匹配算法所确定的位置，然后再进行新位置的模式匹配，直到模式串中所有位都匹配成功，则找到了模式串在主串中的位置，匹配过程成功，在屏幕上输出匹配成功的位置；或者主串中的没有一个子串与模式串能够匹配，则匹配过程失败，在屏幕上输出匹配失败。

3.1. 演示系统整体功能设计

演示系统为使用者提供三种模式匹配算法的演示执行，其中包括 BF 算法模式匹配演示，KMP 算法

模式匹配演示，改进 KMP 算法模式匹配演示。整体系统功能图如图 3 所示。

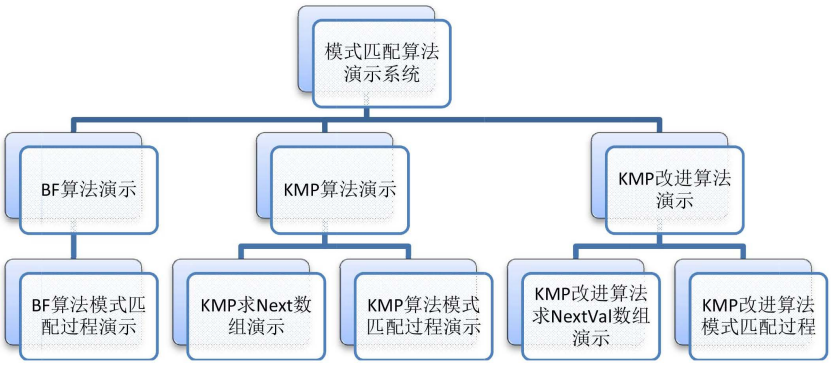

Figure 3. System function diagram
图 3. 系统功能图

3.2. 演示系统中的详细功能

3.2.1. 输入字符串

在进行模式匹配算法的演示执行之前，使用者可以输入任意的、自己指定的主字符串和模式字符串来进行模式匹配算法演示。当使用者输入主串和模式串以后，使用者输入的主串和模式串会经过系统的自动处理，显示在系统的演示界面中，演示界面还包括模式匹配算法的主要代码，以及在模式匹配算法演示执行过程中所生成的中间变量，这些数据都会显示在演示界面中。输入字符串后处理流程如图 4 所示。

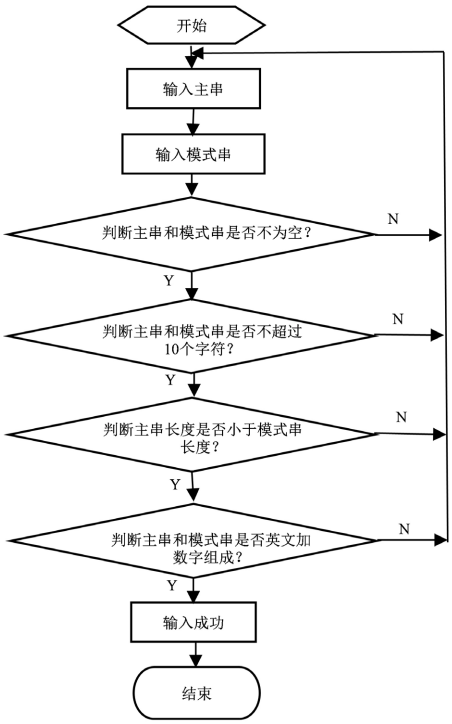

Figure 4. Input string function flowchart
图 4. 输入字符串功能流程

3.2.2. 按钮功能

在演示过程中，使用者可以点击系统提供的相关演示按钮来进行演示操作，这些按钮包括单步按钮，执行按钮，暂停按钮，恢复按钮，修改按钮和返回按钮。使用者可以通过点击单步按钮，来单步执行模式匹配算法中的每一行的代码，根据模式匹配算法中的每一行的代码内容，系统会自动在演示界面中显示数据，并将字符串的模式匹配过程用图形界面编程使其形象化，供使用者学习和理解模式匹配算法的执行原理。点击执行按钮时，系统会自动执行模式匹配演示程序至模式匹配演示程序结束。点击暂停按钮，会在系统自动执行模式匹配演示程序时，终止系统自动执行模式匹配演示程序。点击恢复按钮，系统会恢复到使用者输入主串和模式串时的演示界面，即重新开始演示根据用户输入的主串和模式串的模式匹配程序。以 BF 算法为例的演示主界面如图 5 所示。

从图 5 可以看到，本演示系统还有一个其他演示系统中没有的功能按钮：修改按钮。因为在现实生活中，各种版本的数据结构算法模式匹配代码原理大致相同，但是，代码的中间变量名称却各不相同，为了方便使用者学习模式匹配算法，并且不会因自己所学代码和系统代码中的中间变量名称不同，而搞混淆模式匹配算法原理，系统为使用者提供修改功能，可以对系统代码中的中间变量名称进行修改，类似如主串指针名称和模式串指针名称，可以通过系统提供的修改按钮对主串指针名称和模式串指针名称进行修改，方便使用者对模式匹配算法代码的学习。还有返回按钮，当使用者浏览完演示界面中的模式匹配演示后，使用者想要观看其他算法的模式匹配演示过程，则使用者可以点击返回按钮，返回到可以选择模式匹配算法演示界面，对其他算法的模式匹配演示过程进行浏览观看。

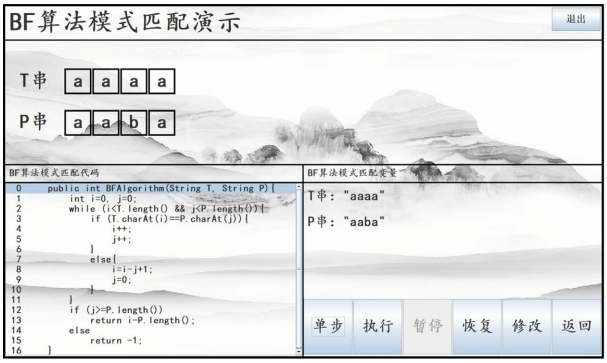

Figure 5. System demonstration main interface (BF algorithm as an example)
图 5. 系统演示主界面(BF 算法为例)

4. KMP 算法的演示过程实现

模式匹配算法中 BF 算法较为简单，这里以 KMP 算法为例介绍一下 KMP 算法的演示过程。

4.1. 求 Next 值的演示

Next 数组值计算如公式(1)所示，实际上这个计算过程就是求模式串自身的模式匹配过程，所以在演示系统中用模式串当主串，模式串也是待匹配的子串模式串，进行从头开始的模式匹配过程，求得模式串中每个字符位置 j 之前的子串中前缀和后缀相同的位数。

演示系统中首先创建继承于 JPanel 类 KMPNextRoot 类用于求 Next 数组演示的根容器，然后在其中设置当前根容器的子控件布局，并设置下一步按钮方法，在其中设置单步按钮事件监听器的具体内容，记录求 Next 数组演示的执行次数，设置界面上具体的显示内容。演示界面如图 6 所示。

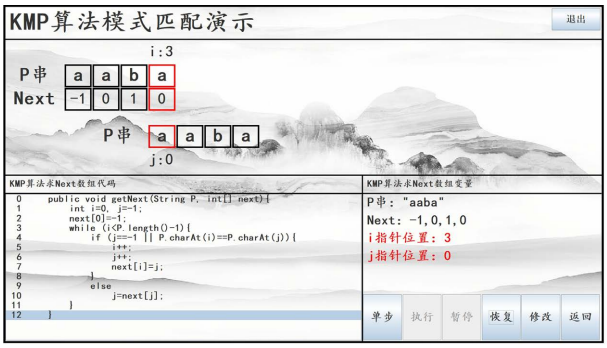

Figure 6. Demonstration interface for finding Next arrays
图 6. 求 Next 数组的演示界面

4.2. KMP 模式匹配过程的演示

求得 Next 数组后，再进行 KMP 无回溯的模式匹配过程，匹配过程参考图 2 所示。实现中首先创建 KMPRoot 类，其继承于 JPanel 类，用于作为 KMP 算法模式匹配演示的根容器。在其构造方法中，设置当前根容器的子控件布局，并设置下一步按钮方法，在其中设置单步按钮事件监听器的具体内容，KMP 算法模式匹配演示的执行次数和突出显示 KMP 算法代码编号，设置界面上具体的显示内容。演示界面如图 7、图 8 所示。

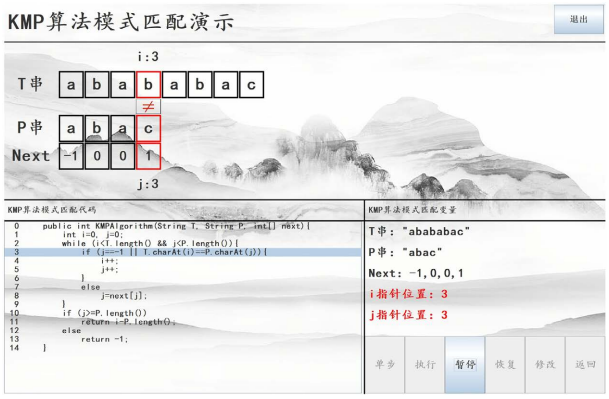

Figure 7. KMP demonstration process interface No. 1
图 7. KMP 演示过程界面 1

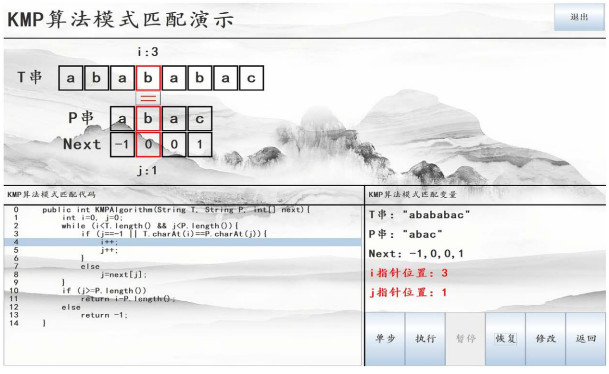

Figure 8. KMP demonstration process interface No. 2
图 8. KMP 演示过程界面 2

改进的 KMP 算法在演示系统中也进行了展示，篇幅原因，论文中不再描述。

在 KMP 算法及改进 KMP 算法的演示系统中，用户可以定制变量名称，以便与书中所讲内容相匹配。比如位置指针名称可以是 i，j，也可以名称是 t，k，由用户自由定制，如图 9 所示。并且在系统中还可以点击“恢复”按钮随时恢复到本次模式匹配最初的状态，方便上课过程中进行不断的重复演示。相对于传统的模式匹配演示系统，如文献[8]中提到的，自由定制指针名称和随时恢复最初状态是本系统为方便用户操作及理解做的一项创新性探索，并且通过实操发现这两项功能非常实用。

Figure 9. Modification of pointer name
图 9. 修改指针名称

5. 总结与展望

在演示系统中，视频、文本、图片、交互动画和演示执行程序，相互配合，从而构建出了一个更有效率的学习环境。本次研究的字符串模式匹配算法演示系统的研究能让使用者通过直观的方法理解字符串模式匹配的原理，了解字符串模式匹配算法的执行过程。后续还会研究开发更多的演示系统，通过演示系统，可以把抽象难懂的数据结构理论数据模型变得容易理解，有助于更好地学习和运用数据结构中的内容，提高了数据分析计算的效率与效果。

基金项目

北京市人文社科基金重点项目：大数据背景下北京市属高校新工科人才培养质量过程性评价研究(19JYA001)资助。教学改革项目：促进高校分类发展－专业建设－计算机学院专业建设与人才培养模式改革资助。

参考文献

[1] 谭永滨, 侯梦飞, 张志军, 李小龙, 程朋根, 章泽之. 基于模式匹配的交通微博文本位置信息提取模型[J]. 地理与地理信息科学, 2021, 37(5): 16-22.

[2] 徐琳, 魏晓超, 蔡国鹏, 王皓, 郑志华. 一个高效的安全两方近似模式匹配协议[J]. 计算机研究与发展, 2022, 59(8): 1819-1830.

[3] 曹丽娜, 王霞, 周瑛. 基于模式匹配算法的空间属性数据挖掘仿真[J]. 计算机仿真, 2022, 39(9): 273-276.

[4] 李一鸣, 刘胜利. 自适应安全的支持模式匹配的流加密方案[J/OL]. 西安电子科技大学学报: 1-13. http://kns.cnki.net/kcms/detail/61.1076.tn.20230307.1635.002.html, 2023-06-29.

[5] 周琰, 马强. 基于混合模式匹配算法的网络入侵检测[J]. 计算机测量与控制, 2022, 30(11): 65-70.

[6] 王乂冉. 基于近似模式匹配的并行压缩算法的研究与实现[D]: [硕士学位论文]. 长春: 吉林大学, 2020.

[7] 刘邦国, 陈庆春, 类先富. 一种面向 PDF 文本内容审查的高效多模式匹配算法[J]. 计算机应用研究, 2020, 37(6): 1755-1759.

[8] 王宏, 曹家庆, 黄斌, 陈琪. 基于 Java 的数据结构算法演示系统[J]. 南昌航空工业学院学报(自然科学版), 2006, 20(2): 70-75.

[9] 王玢玥, 李冬梅, 李华颖, 姚佳璐, 王仁生. 数据结构算法演示系统的设计[J]. 教育教学论坛, 2016(28): 167-168.

[10] 万东洋. 基于 Java 的数据结构算法演示系统研究[J]. 内蒙古煤炭经济, 2019(20): 171-172.

[11] Knuth, D.E., Morris, J.H. and Pratt, V.R. (1977) Fast Pattern Matching in Strings. *Siam Journal on Computing*, **6**, 323-350. https://doi.org/10.1137/0206024