

快速模式匹配在网络教育平台中的应用研究

刘 啸

(徐州师范大学 现代教育技术中心, 江苏 徐州 221116)

摘 要:网络教育平台建设的核心是教学资源库的建设,各种类型资源的快速匹配是需要频繁使用的基础功能。现有的解决方案往往使用的是BF等经典算法,匹配效率有一定的提高空间。用XML技术作为平台中各种资源描述的统一工具,通过对XML技术中DTD模式匹配方法的研究以及对模式匹配快速算法的改进,提高模式匹配的效率,从而设计出一种有效的信息模式匹配技术模型。通过仿真实验表明该模型可有效地提高网络教育平台中信息匹配的效率,进而为平台中资源的整合与模式的统一提供支持。

关键词:网络教育平台;DTD模式;模式匹配快速算法

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2012)03-0210-04

Application Research of Fast String-Matching in Network Education Platform

LIU Xiao

(Modern Educational Technology Center of Xuzhou Normal University, Xuzhou 221116, China)

Abstract:Teaching resource library construction is the core of online education platform, various types of resources fast matching is frequently used basis functions. Existing solutions tend to use the BF and other classic algorithms, matching efficiency has a certain room for improvement. Using XML technology as a unified description of various resource tools in a platform, researching DTD pattern matching method in XML technology and improving fast algorithm for pattern matching, improve the efficiency of pattern matching to design an effective information pattern matching technology model. Simulation results show that the model can effectively improve the efficiency of information matching in online education platform, and thus in a platform provide support for the integration of resources and the unified model.

Key words:network education platform; DTD; fast string-matching algorithm

0 引 言

随着信息技术和网络技术的发展和基于互联网的在线教育系统的普及,人们获取新知识、新信息的方式日新月异。网络教育克服了传统教育在时间、空间上的限制,作为一种崭新的远程教育方式正逐渐运用到实际的教学活动中。网络教育对于有效地发挥现有的各种教育资源优势、实现资源的合理配置起着非常重要的作用。

网络教育的优势主要体现在双向交流的便捷和深度上,通过现代信息技术实现同步和异步通信,使教与学更加开放、灵活,更具多样化和个性化,学生之间、师生之间的交流得到进一步的加强,信息交换的种类和

容量都显著增加,交流形式更趋加多样化^[1]。师生之间、学生之间的交流不再受时间和空间的限制,可以通过电子邮件、聊天室和电子公告牌等进行。不仅如此,交流的速度和质量也得以大大提高,交流的内容也更加复杂与广泛。网络教育平台的基本功能结构如图1所示。

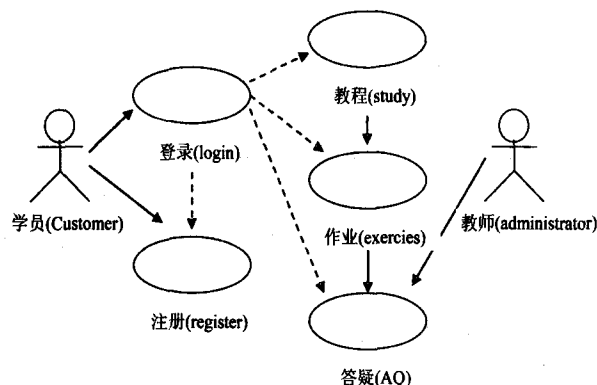


图1 网络教育平台基本功能结构

网络平台的建设、课程集合的建设、电子教学资源

收稿日期:2011-08-05;修回日期:2011-11-09

基金项目:江苏省现代教育技术研究2010年度立项课题(2010-R-15841);徐州师范大学科研基金项目(09XLB23)

作者简介:刘 啸(1979-),男,江苏徐州人,讲师,硕士,研究方向为信息安全、网络与数据库。

素材库及其管理系统的建设是网络教育平台建设的核心^[2]。在以上应用系统的开发过程中,信息的快速匹配是频繁需要使用到的基础功能,例如标准格式的教学素材的收集与整合、试题库的创建、网络考试模块的试题批改功能等。

现有的网络教育平台在涉及到信息匹配时,大多使用 BF、KMP 等经典算法,在信息查找效率和速度上都有一定的提高空间,因而需要引入更加有效的快速模式匹配算法。

1 网络教学平台中 XML 数据与 DTD 模式匹配

1.1 XML 简介

XML 属于 SGML (Standard Generalized Markup Language, 标准的通用标记语言) 的一个子集合,它实现了 SGML 的大部分功能,但是降低了它的复杂性^[3]。XML 由 W3C 下属支持的 XML 工作小组在 1996 年开发完成,1998 年 2 月正式成为一种 W3C 推荐标准。

XML 被设计成 SGML 的简化子集,用于描述和操纵具有较短生命周期的文档,在不同的平台间作为信息的共用模式进行信息传递是其强项,而且特别适合于跨操作系统平台的 WEB 环境。

1.2 XML 与 HTML

人们经常会将 HTML 和 XML 进行比较,并常常用 XML 来代替 HTML,那么这两种标记语言有什么相同和不同之处呢。这两种规范都直接使用文本文件作为它们的内容的载体,都使用尖括号来包含 SGML 风格的标记。但是,HTML 仅具有实现设计好功能的一系列标签符号,可以使用它们来修饰自己的内容,而 XML 允许用户自己定义一组全新的标签符号及其相互关系^[4]。可以说 XML 是灵活度更高、可创造性更强的 HTML。

1.3 用 XML 作为平台中各类资源描述的标准

适应性很强的数据存储格式、标记符号系统的可扩展性、高度结构化和网络传输应用的密切联系是 XML 体系的主要特征。XML 作为网络教育平台中各种教学资源的描述、封装、管理和应用开发方面有几个优势。

(1) XML 具有良好的语义和清晰的结构,又可以根据具体情况自定义所需的标记;XML 又是网络上不同操作系统平台间进行数据交换的理想载体。所以它可以对各种教学资源(课件、试卷库、教案、素材库等)进行标准化的格式设定。

(2) XML 把数据的数值内容与其表达形式分离开来,它的侧重点在于描述数据及其之间的关系。XSL 和 CSS 用来对数据内容进行修饰和包装,它们是将特

定的样式信息传递给命名的 XML 元素的基本手段,这就像所熟悉的 WORD 字处理程序一样,更改文档中的样式信息(字体、颜色、倾斜度等等)。实现了形式与内容的分离。

(3) 常见的教学资源的组织结构是树状结构。例如讲稿课件就是按章节进行组织,有时需要对其某些章节的顺序进行调整,也就是说需要对这些资源根据实际情况动态组织。XML 的文档组织结构正好与之相符合,利用 XML 中的 Xpath 技术可以表达教育信息资源间复杂的关系,创建其灵活的组织结构^[5]。

1.4 基于 DTD 的数据模式匹配

XML 有一套预定义的语义规则,具体规范包括 DTD (Document Type Definition) 和 Schema 两种。而且,像传统的关系型数据库一样,XML 的语义规则可以用来进行 XML 文档格式的验证和 XML 文档索引的检索。DTD 作为一种对 XML 数据模式进行定义的方法有以下功能。首先,提供具有约束力的 XML 数据结构,通过检查 DTD 结构定义,系统可以检测 XML 文档是否符合其 DTD 规范。其次,使用的 DTD 是没有严格的推荐标准的,其具有一定的包容性,允许 XML 数据在一定程度上不符合 DTD 的规范定义。

网络教育平台中的各种资源通过 XML 进行描述后,由于数据交换的需要,亟需一种手段来对这些格式并不十分严谨但却在很大程度上遵守 DTD 模式要求的不同 XML 数据进行相似度匹配,也就是 XML 数据的模式匹配^[6]。

预先设计好网络教育平台中各种资源统一使用的标准 XML 数据交换格式,并称之为源 DTD 模式,把平台的使用者所发布的特定 XML 数据格式称为目标模式。

计算目标模式与源 DTD 模式之间的相似度,可以采取以下办法:将 DTD 看成是约束 XML 数据的一组规则集合,利用每个 XML 数据与 DTD 模式进行相似度的比较计算。该方法需要研究和设计一种有效的模式匹配方法。

2 使用模式匹配快速算法

2.1 提高模式匹配算法性能的基本思路

经典的字符串模式匹配算法的思路是:主串和模式串都自左向右,每个字符依次进行比较,如果出现不匹配,用来进行主串和模式串的位置标记的指针都要向左回溯^[7],当查找空间规模较大时查找时间较长。

在定义模式串的特征值后,改进过的模式匹配快速算法通过判断匹配必要条件以及两相邻字串的特征值之间的递推关系,找出其中规律尽量避免指针回溯。改进算法能够大大提高模式匹配的效率,从而能够显

著提高网络教育平台中的信息匹配效率。

算法的改进思路是:设主串 S 的长度为 n , 模式串 T 的长度为 l (设定 l 远小于 n), 对主串 S 的结构进行以下步骤的分析。

(1) 把主串按照匹配的要求分解成若干子串 Sk 进行进一步判断, Sk 的个数为 $n-l+1$ 。 Sk 中能与 T 匹配的子串, 设为 m 个。一般来讲 m 是一个远小于 $n-l+1$ 的数。

(2) 找出所有与 T 匹配的子串 Sk 和找出所有与 T 不匹配的子串 $Sk^{[8]}$ 是模式匹配算法的最终目标。提高整体算法效率的侧重点应放在不匹配子串 Sk 的判断效率上。

(3) 根据两串匹配的必要条件, 如果能够按某种固定模式求出每一个子串 Sk 的特征值, 只要与用相同固定模式求出的 T 的特征值进行比较, 即可判断 Sk 与 T 是否不匹配。

(4) 求解特征值的关键应放在找寻相邻子串间的联系上, 通过两相邻子串 Sk 和 $Sk+1$ 之间的相互联系, 顺次求出所有长度为 l 的子串的特征值^[7]。 $Sk+1$ 可由 Sk 去掉首字符的其余部分链接字符 $Sk+l$ 后得到, 如图 2 所示。



图 2 相邻子串间的关系

(5) 因此得到两相邻子串 Sk 和 $Sk+1$ 的对应的特征值 $V_{s_{k+1}}$ 与 V_{s_k} 之间有如下关系:

$$V_{s_{k+1}} = V_{s_k} - \text{code}(sk) + \text{code}(sk+l).$$

表达式中的 $\text{code}(c)$ 表示对于字符 c 的某种编码过程。

(6) 采用三点匹配算法通过特征值的求解尽快检测出不匹配子串 Sk 。三点匹配算法的思路是依次检测首字符、尾字符和中字符是否匹配, 从而得出结论。如此处理的目的是扩大检测范围^[9]。

2.2 快速模式匹配改进算法

通过以上对于算法改进思路的描述, 给出判断特征值相等的子串是否与模式串匹配的三点匹配算法^[10], 以及关于主串的模式匹配快速算法。

三点匹配算法:

If (首字符 t_1 和 s_k 相等)

{ If (尾字符 t_l 和 s_{k+l-1} 相等)

{ If (中点字符 $t_{[(l+1)/2]}$ 和 $s_{[k+(l-1)/2]}$ 相等)

{ If (左子串 $t_2t_3 \dots t_{[(l+1)/2-1]}$ 和左子串的 $s_{k+1}s_{k+2} \dots s_{[k+(l-1)/2-1]}$ 匹配)

{ If (右子串 $t_{[(l+1)/2+1]}t_{[(l+1)/2+2]} \dots t_{l-1}$ 和右子串的

$s_{[k+(l-1)/2+1]}s_{[k+(l-1)/2+2]} \dots s_{k+l-2}$ 匹配)

{ Sk 与 T 匹配 }

Else { Sk 与 T 不匹配 } }

Else { Sk 与 T 不匹配 } }

Else { Sk 与 T 不匹配 } }

Else { Sk 与 T 不匹配 } }

Else { Sk 与 T 不匹配 }

模式匹配快速算法:

(1) 用以下的表达式求解模式串 T 的特征值^[7]:

$$V_T = \sum_{i=1}^l \text{code}(T[i]), \text{主串 } S \text{ 长度为 } l \text{ 的子串 } S_1 \text{ 的特征}$$

$$\text{值: } V_s = \sum_{i=1}^l \text{code}(S[i]).$$

(2) for ($k=1$; $k \leq n-l$; $k++$) {

// 按照自左向右的顺序判定子串 Sk 是否匹配。

if ($V_s \neq V_T$) {

// 如果不匹配, 子串起始位置右移一位, 计算新子串的特征值。

$$V_s = V_s - \text{code}(S[k]) + \text{code}(S[k+l])$$

} }

else {

调用三点匹配算法

if (不匹配) {

$$V_s = V_s - \text{code}(S[k]) + \text{code}(S[k+l])$$

else { 输出 k

$$V_s = V_s - \text{code}(S[k]) + \text{code}(S[k+l])$$

} }

(3) 调用三点匹配算法判断子串 S_{n-l+1} 是否与模式串 T 匹配。

(4) 算法结束。

3 实验结果

3.1 实验描述

为验证改进算法的实际效果, 实验选择 JBuilder 开发平台, 使用 Java 语言实现改进过的模式匹配快速算法。并以 BF 经典算法作为参照, 对网络教育平台中的模式匹配问题进行单机仿真^[11]。计算机配置: CPU: Intel® Core2 Duo 3.0GHz; 内存: 2GB。

在实验中, 设置的场景为网络教育平台素材库中, 各种形式素材的 XML 描述与相应的 DTD 模式进行模式匹配。实验中根据素材 XML 描述字串的长度不同选择 1000 组样本, 并将其字串长度控制在 1kB 至 100kB 范围内, 对改进的模式匹配快速算法和 BF 经典算法进行匹配时间比较 (见表 1)。通过这些仿真实验数据, 可以验证所提出的改进算法在执行效率上有所提高。

3.2 结果分析

从图 3 可以看出, 随着匹配字串长度的递增两种

算法的字串匹配时间都在上升,但 BF 算法上升的趋势更为明显。所以,文中提出的改进算法比 BF 算法更具优势,性能也更加优越,同时,也充分说明了该算法的有效性和实用性。

表 1 标志性长度字串匹配速度

字串长度(kB)	BF(ms)	改进算法(ms)	字串长度(kB)	BF(ms)	改进算法(ms)
10	1.667	1.565	60	11.044	9.613
20	3.512	2.981	70	13.491	10.940
30	5.236	4.510	80	15.742	13.016
40	6.996	6.291	90	17.477	14.837
50	8.830	8.124	100	19.576	16.295

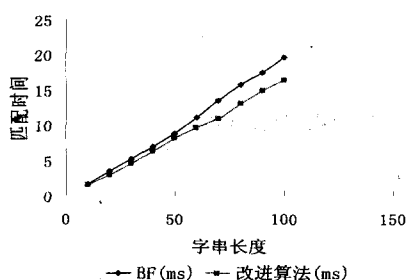


图 3 改进算法字串匹配效率比较

4 结束语

通过实验验证,改进过的模式匹配快速算法可以提高本系统中目标模式与源 DTD 模式的匹配效率,进而为平台中资源的整合与模式的统一提供支持。在今后的研究工作中可以进一步设计计算目标模式与源

DTD 模式之间的相似度的模型^[12],探讨二者相似度与改进的模式匹配快速算法匹配效率之间的关系。

参考文献:

- [1] 张维蔚,李 超.一种多策略 GML 应用模式匹配方法[J].北京航空航天大学学报,2008(5):520-523.
- [2] 王晓东.计算机算法设计与分析[M].北京:电子工业出版社,2004.
- [3] 陶世群,富丽贞.一种高效非归并的 XML 小枝模式匹配算法[J].软件学报,2009(4):795-802.
- [4] Deitel H M, Deitel P J. XML 编程技术大全[M].美国:美国培生出版集团,2006.
- [5] White C. XSLT 从入门到精通[M].北京:电子工业出版社,2002.
- [6] McLaughlin B. Java & XML[M]. 2nd ed. [s. l.]: O'Reilly & Associates, 2005.
- [7] 刘玉龙,刘 啸.一种模式匹配快速算法[J].计算机科学,2008,35(1):219-220.
- [8] 陶善旗,李 俊,郭伟群,等.入侵检测系统中模式匹配算法的研究与改进[J].计算机技术与发展,2010,20(2):167-174.
- [9] 林倩瑜,冯少荣,张东.基于神经网络和模式匹配的股票预测研究[J].计算机技术与发展,2010,20(5):17-25.
- [10] 程玉青,梅登华.入侵检测系统中 BM 模式匹配算法的改进[J].计算机技术与发展,2009,19(3):172-174.
- [11] 蒋 科,郑有才.一种高效的 XMLQuery 基本模式匹配算法[J].计算机技术与发展,2007,17(7):87-90.
- [12] 甘学士,孙力娟.改进的模式匹配算法及在入侵检测中的应用[J].计算机技术与发展,2006,16(7):150-152.

(上接第 209 页)

4 结束语

在小车制作完成后,进行了软硬件的调试,表明该小车不但能够通过无线通讯来控制小车的行驶,还能够让小车自主行驶,实现了智能化,性能稳定,可改装成智能机器人用于工业生产等实践工程中,当然部分设计仍需修改,如对 SPWM 可用硬件来实现,从而更适合于实时控制。

参考文献:

- [1] 陆凯峰.基于 DSP 的多功能智能机器小车的研究[D].无锡:江南大学,2008.
- [2] 谢宝昌,任永德.电机的 DSP 控制技术及其应用[M].北京:北京航空航天大学出版社,2005.
- [3] 王亚嫔.基于 DSP 的二相混合式步进电机多细分驱动器的研究[D].杭州:浙江工业大学,2008.
- [4] 王晓明,王 玲.电动机的 DSP 控制—TI 公司 DSP 应用

- [M].北京:北京航空航天大学出版社,2004.
- [5] 关保国,钟伟弘.步进电机的驱动及微控制[J].自动化与仪表,2000,15(5):55-56.
- [6] 张丹红,游珍珍.DSP 的多领域应用研究[J].计算机技术与发展,2006,16(3):206-207.
- [7] 张卫宁.TMS320C28X 系列 DSP 的 CPU 与外设[M].北京:清华大学出版社,2005.
- [8] TMS320x28xx,28xxx DSP Peripherals Reference Guide[M]. [s. l.]: [s. n.], 2009.
- [9] TMS320F2809, F2808, F2806, F2802, F2801, C2802, C2801, and F2801x DSPs Data Manual[M]. [s. l.]: [s. n.], 2009.
- [10] 卜雪民,钱 峰.基于 DSP 的智能车辆系统移动小车的设计[J].电子测量技术,2007,30(10):80-83.
- [11] 吴 莹,陈延明,沈祺钢.基于 DSP 的 SPWM 波设计与实现[J].微计算机信息,2008,24(32):187-189.
- [12] Lin Tieluo, Zhang Jianxun. DSP-based microstep controller of stepper motor[C]//Fifth World Congress on Intelligent Control and Automation. [s. l.]: [s. n.], 2004:4441-4445.