



(12) 发明专利

(10) 授权公告号 CN 101901257 B

(45) 授权公告日 2012.07.04

(21) 申请号 201010232463.2

审查员 杨春雨

(22) 申请日 2010.07.21

(73) 专利权人 北京理工大学

地址 100081 北京市海淀区中关村南大街 5 号

(72) 发明人 嵩天 黎达

(51) Int. Cl.

G06F 17/30(2006.01)

(56) 对比文件

CN 101051321 A, 2007.10.10, 全文.

CN 101251845 A, 2008.08.27, 全文.

CN 1477559 A, 2004.02.25, 全文.

Tian Song 等. AN ALGORITHM OF

LARGE-SCALE APPROXIMATE MULTIPLE STRING MATCHING. 《Department of Computer Science and Technology》. 2006, 全文.

权利要求书 3 页 说明书 10 页

(54) 发明名称

一种搜索引擎中的多字符串匹配方法

(57) 摘要

本发明涉及一种多字符串匹配方法,属于字符串匹配技术领域。本发明基于传统的 Wu-Manber 方法,对规则集中的长字符串和短字符串进行了分离,在建 SHIFT 表时,对规则集中的长字符串和短字符串采取不同的处理方式,从而使得 SHIFT 表的最大表项不受到短字符串长度的限制,克服了其最大跳跃距离受到规则集中最短字符串长度限制的不足;通过引入 HOT 表,利用匹配过程中查找 HOT 的方法,使得窗口的最大跳跃距离得到增大的同时不会遗漏短字符串。本发明方法具有更高的匹配效率。