

基于Trie树的快速电子词典开发

蔡 琪, 刘东霞

(平顶山教育学院, 河南 平顶山 467000)

摘 要: 当今社会在步入一个大数据时代, 时间和效率举足轻重。因此设计和开发出一款能快速检索目标词汇的电子词典具有十分重要的现实意义。开发的电子词典系统运用 Windows API 开发, 采用 Trie 树的数据结构设计。结果表明, 电子词典实现了 Trie 树结构的存取和快速 Hash 映射查询, 实现主流电子词典常用功能, 包括单词查找、添加生词、我的单词本、课程设置、单词测试和帮助等, 可满足大部分用户的需求, 具有良好的扩展性。

关键词: 快速检索; Trie 树; Hash 查找; 电子词典

中图分类号: TN919-34; TP393

文献标识码: A

文章编号: 1004-373X(2014)12-0090-03

Development of rapid electronic dictionary based on Trie tree

CAI Qi, LIU Dong-xia

(Department of Computer Science, Pingdingshan Institute of Education, Pingdingshan 467000, China)

Abstract: As a foreign language plays a more and more important role, a electronic dictionary for quick retrieval target vocabulary was designed and developed. The Windows API and Trie tree data structure are adopted in the electronic dictionary system design and development. The access and rapid HASH map check word of Trie tree structure were realized in the electronic dictionary. The common functions of the mainstream electronic dictionary, including word lookup, new words addition, my words book, curriculum setting, word test and help, were implemented. It can meet the needs of most users and has good scalability.

Keywords: rapid retrieval; Trie tree; Hash lookup; electronic dictionary

随着外语发挥着越来越重要的作用,与此同时,社会正在步入一个大数据时代,时间和效率举足轻重^[1]。通过深入学习和研究程序设计技术、数据库系统开发和应用,设计和开发出一款能够满足不同用户需求的多功能电子词典系统,以帮助英语学习者更方便、更快捷地查询单词、记忆单词,既有效,又自由地对词库进行管理和操作^[2-3]。采用 Trie 树结构构造单词查找树和快速 Hash 映射取词,可在占用少量计算机硬件资源的基础上,实现快速查询目标词汇的功能。因此设计和开发这款能快速检索目标词汇的电子词典有十分重要的现实意义,能很好应对在大数据时代下有效地满足了用户需求。

1 Trie 树简介

1.1 Trie 树简介

Trie 树又称为字典树,是一种树形结构,它是一种哈希树的变种,是一种用于快速检索的多叉树数据结构^[4]。典型应用是用于统计和排序、查询大量的字符串(但不限于字符串),所以经常被搜索引擎系统用于文本的

词频统计等。Trie 树也有其缺点, Trie 树的内存消耗非常大。当然,用左儿子右兄弟的方法建树,可能效果会好一些^[5]。

1.2 Trie 性质

很多人认为 Trie 的根节点不包含任何字符信息,在此认为习惯的 Trie 根节点却包含信息,而这样也方便,下面叙述 Trie 的性质(基于本文所讨论的简单 Trie 树)^[6-7]:

(1) 字符的种数决定每个节点的出度,即 branch 数组(空间换时间思想)。

(2) branch 数组的下标代表字符相对于 a 的相对位置。

(3) 采用标记的方法确定是否为字符串。

(4) 插入、查找的复杂度均为 $O(\text{len})$, len 为字符串长度。

1.3 Trie 的示意图

如图 1 所示,该 Trie 树存有 abc, d, da, dda 四个字符串,如果是字符串会在节点的尾部进行标记^[8-10]。没有后续字符的 branch 分支指向 NULL。

2 设计与实现

2.1 系统设计思想

电子词典软件面向用户时,重要的是其查询效率与

收稿日期: 2014-03-15

基金项目: 河南省科技计划重点项目(102102210416)

可信性,即用户能迅速而又准确地查询到词语的相关注释。设计本电子词典主要是为了用于帮助用户查找一些不懂的单词及其相关内容。故本系统应该支持以下电子词典的核心功能:

- (1) 查询单词功能。具有很高的查询效率。
- (2) 单词测试功能。分为知道英语单词选择中文解释和知道中文解释选择英语单词两项功能。
- (3) 生词添加和新词构造功能。分为单个添加和批量添加功能。
- (4) 词汇查看功能。核对和查看各词库词汇及信息。
- (5) 我的生词本功能。自定义的词库及单词测试功能。
- (6) 课程设置功能。

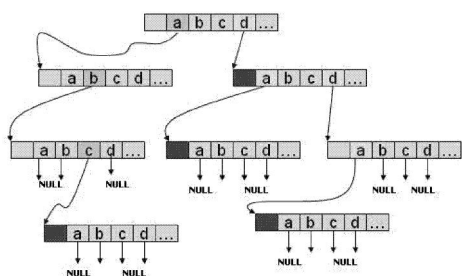
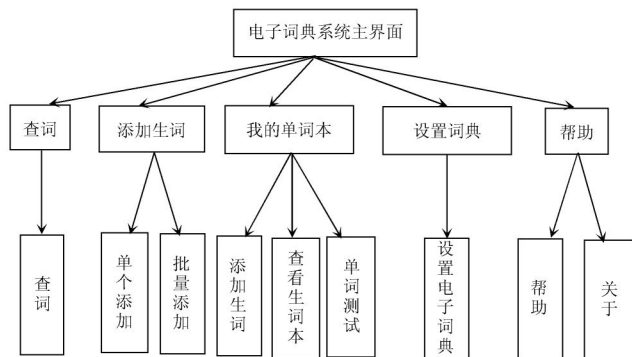


图1 Trie树结构示意图

2.2 系统功能结构问题

系统功能模块图如图2所示。



2.2.1 单词查询

单词查询是电子词典最基本也是最重要的功能,只要用户输入想要查询的单词,电子词典将自动帮助用户进行单词查询,当词库中有该词汇,电子词典将快速返回词库中关于该词汇的相关信息给用户。当用户输入时,应用程序能够智能列出所有相近备选的单字,这可以方便用户在没有完成整个单词输入的情况下查询出目标词汇结果。返回结果包括所查询单词的记性、音标和中文翻义等。同时,在电子词典显示板下方会智能地显

示部分相近的词汇的查询结果。实现方法:用户敲击键盘输入某字符时,操作系统会发出一个WM_KEYDOWN的消息,然后电子词典应用程序有相应的回调函数自动响应对所构造的单词查找树进行前序遍历,遍历到当前输入的字符,得出查询结果显示在电子词典显示板上。然后把部分相近的单词也绑定在结果显示板上。

2.2.2 添加生词

生词添加分为单个添加和批量添加。单个单词添加是在用户界面上输入要添加的新单词及单词相关的信息,然后保存到指定的生词库即可。而批量添加则是在用户界面上选择要添加单词库文件的路径,单词库文件须事先按要求及格式准备好,然后点击确定添加即可。

2.2.3 我的单词本

我的单词本包括以下几个功能:添加词汇功能(注意:此处的添加词汇功能实现与上面的添加词汇功能一致,只是所添加的词汇是添加到用户自定义生词本上);单词查看功能;单词测试功能。

(1) 添加生词:与上面添加词库中的单个添加功能一致,只是重写向的文件为用户自定义生词本。

(2) 词汇查看。电子词典词库存在出现遗漏、错误的可能性,也可能词库单词不够完善的情况。相应的电子词典系统应该有一种可让用户查看词库内容的功能,因此把电子词典设计成能随时让用户查看任意词库的所有单词和信息,以满足用户要添加或修改词库的意愿。词汇查看功能主要就是选择相应的词库,然后把词库中的单词及信息呈现在用户界面上,用户可以按序地去核对词库,查看词库是否存在遗漏,或发生错误,或缺少一些词汇。用户也可以通过查询来检查该单词是否存在生词库中。若发现词库存在遗漏,则可以及时添加;若词库中单词发生错误,则可以及时个性;若发现一些词汇不在词库中,则可以通过上述的添加功能添加进指定词库中。

(3) 单词测试。单词测试功能,能以游戏选择的形式达到让读者轻松背诵或复习单词的目的,该电子词典针对不同用户的不同需求可选择不同生词库来进行单词测试,从而让用户对自己的词汇水平有一个大体的了解,同时用户也达到了背诵和复习单词的目的。

2.2.4 课程设置功能

在用户未自定义词库之前,本电子词典系统已经包括有考研词库、四级词库、六级词库、生词库、GRE的几大词库。加上用户自定义的词库,必然会让词库变得更大,然后针对不同的用户对象,会有不同的用户需求,因此,在电子词典系统进设计具有选择词库,设置词库功能,可方便不同的用户,提高用户的学习效率,增强电子

词典软件的友好度。

在初次设置课程时,程序会在本地的电子词典的目录下生成一个名词 config.ini 的文件,里面存放用户的课程设置情况。当初次运行时,在未设置课程前则系统无法进行单词测试功能。在要重新选择课程时,电子词典应用程序会对该文件作出修改,修改成为当前所选的词典,以便于满足不同的用户需求。

在用户要进行单词测试或查看词库信息时,首先会到磁盘中找本电子词典项目目录下是否有 config.ini 这个文件。若没有发现此文件,则要求用户选设置课程,并将用户所选课程信息生成 config.ini 文件,然后在该课程下进行单词测试或查看词汇信息即可。若在项目目录下发现有此配置信息文件,直接从 config.ini 文件中读取课程信息,然后以该信息进行单词测试或查看词汇信息即可。

2.2.5 帮助功能

作为一款面向用户的应用软件,提供软件的操作说明和帮助是十分有必要的,是衡量软件友好度的一个重要指标。当用户初次接触此软件时,因为有了操作说明或者是帮助,能帮助他们快速上手,熟悉软件界面,尝试运用软件实现功能达到合用软件的目的。倘若没有操作说明和帮助,那么用户只能通过自己的摸索来对软件进行操作。在摸索过程中,许多用户因为缺乏耐心或软件操作过于复杂而放弃使用此款软件,这是软件开发的一个大忌,无疑对软件的推广产生不良影响。因而,软件提供一定的操作说明或帮助,能让用户用得方便舒服,软件效益也能随之得到提升。

3 结 语

本文论述了一个基于 Trie 树的快速电子词典的设

计与实现,展现了本电子词典在大数据时代中所能发挥出的优点。主要是采用基于 Trie 树为数据结构,以本地文件为词库以及用 Windows API 进行编码实现的电子词典,通过程序构建本地文件流对象,将词库单词读入内存,并在内存中构建一棵字典查找树,这样使得单词查询只需在内存中进行,有效提高了查询目标词汇速度。

参 考 文 献

- [1] KNOWLES F E. The Computer in lexicography [M]. [S.l.] : Cit-eULike , 1990 : 1645-1672.
- [2] ZHANG Yi-hua. Think for the development strategy of the electronic dictionary in our country [J]. Lexicographical Study , 2007(2) : 112-119.
- [3] DE SCHRYVER Gilles-Maurice. Lexicographers' dreams in the electronic-dictionary age [J]. International Journal of Lexicography , 2003 , 16(2) : 143-199.
- [4] BOGURAEV , B , BRISCO T. Computational lexicography for natural language processing [M]. [S.l.] : Longman , 1989.
- [5] PUSTEJOVSKY J , BERGLER S. Lexical semantics and knowledge representation [J]. Lecture Notes in Artificial Intelligence , 1992 , 627 : 86-95.
- [6] YOKOI Toshio. The EDR electronic dictionary [J]. Communications of the ACM , 1995 , 38(11) : 42-44.
- [7] 王博文.通用类 Trie 树及自动生成[J].计算机应用 , 2000(12) : 74-75.
- [8] 朱文强.Trie 树和单字倒排相结合的汉英词典查找机制[J].哈尔滨工业大学学报 : 自然科学版 , 2008(2) : 182-185.
- [9] 何鸿君.一种简单高效的电子词典组织策略[J].计算机科学 , 1996 , 23(2) : 56-57.
- [10] 李娜.用于文本智能处理的电子词典的一种设计方法[J].南京师范大学学报 , 2003(3) : 31-35.

作者简介:蔡 琪(1966)女,河南平顶山人,讲师。主要从事计算机技术应用方面的研究工作。

刘东霞(1979)女,河南平顶山人,讲师。主要从事计算机技术应用方面的研究工作。

(上接第89页)

- [3] 周斌,刘涛,文俊武.GIS 技术在地震学研究中的应用[J].地球物理学进展 , 2011(1) : 685-688.
- [4] 周斌.基于 GIS 的防震减灾计算机信息管理及辅助决策系统 [J].西北地震学报 , 2010 , 23(1) : 1-8.
- [5] 谢建华,陶红,李培铮.开发 WebGIS 的一种新技术 : ArcIMS 技术[J].自然杂志 , 2012(9) : 57-60.
- [6] 高晓红.基于 GIS 的城市防震减灾信息系统研究[J].自然灾害

学报 , 2009 , 12(4) : 110-114.

- [7] 毛继国.基于 WebGIS 的地质灾害预警预报信息系统的设计与实现[J].防灾技术高等专科学校学报 , 2010 , 8(2) : 73-76.
- [8] 郭伦,张晶,唐大仕,等.基于 WebGIS 的体系结构研究[J].地理学与国土研究 , 2011 , 11(4) : 20-24.
- [9] 刘伟庆,徐敬海.基于 GIS 的城市防震减灾信息系统开发[J].南京工业大学学报 , 2011 , 23(5) : 58-65.

作者简介:周昌贤(1981)男,福建厦门人,工程师。研究方向为防震减灾。

郑韶鹏,高级工程师。研究方向为防震减灾。