

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 17/30 (2006.01)

G06F 12/02 (2006.01)



# [12] 发明专利说明书

专利号 ZL 200610078690.8

[45] 授权公告日 2009年4月8日

[11] 授权公告号 CN 100476821C

[22] 申请日 2006.4.30

[21] 申请号 200610078690.8

[73] 专利权人 北京金山软件有限公司

地址 100083 北京市海淀区北四环中路  
238号柏彦大厦20层

[72] 发明人 廖 晔 李成祥

[56] 参考文献

CN1041356C 1998.12.23

CN1647072A 2005.7.27

CN1152326C 2004.6.2

US6507846B1 2003.1.14

审查员 庄锦军

[74] 专利代理机构 北京集佳知识产权代理有限公司

代理人 逯长明

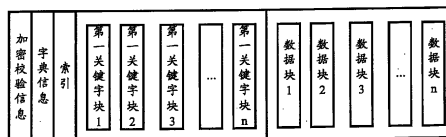
权利要求书3页 说明书11页 附图5页

[54] 发明名称

一种数据存储/检索方法及系统

[57] 摘要

本发明公开了一种数据存储/检索方法及系统。所述存储方法为：提取关键字的字符，将所述提取的关键字的字符作为索引进行存储；存储由首字符相同的关键字组成的第一关键字块；分块存储第一关键字块对应的数据。所述检索方法包括：1) 查找索引，根据索引所标识的第一关键字块和对应数据块的存储位置获取存储的与被检索关键字首字符相同的第一关键字块和对应数据块，所述索引由提取的关键字的字符组成，所述第一关键字块由首字符相同的关键字组成，所述对应数据块根据所述第一关键字块进行存储；2) 查找第一关键字块中含有的被检索关键字所对应数据的存储位置，依据该位置从数据块中获取数据。本发明可提高无线设备检索词典类数据的速度。



1、一种数据存储方法，其特征在于：

提取关键字的字符，将所述提取的关键字的字符作为索引进行存储，所述索引标识第一关键字块及其对应数据的存储位置；

存储由首字符相同的关键字组成的第一关键字块；

分块存储所述第一关键字块对应的数据。

2、如权利要求1所述的数据存储方法，其特征在于，所述索引通过第一关键字块标识元组标识：

所述第一关键字块的首字符；

所述第一关键字块的存储位置；

所述第一关键字块对应的数据块的存储位置。

3、如权利要求2所述的数据存储方法，其特征在于：

进一步存储由第一关键字块中前两位字符相同的关键字组成的第二关键字块；

所述索引进一步通过第二关键字块标识元组标识第二关键字块的前两位字符组合、第二关键字块的存储位置；

所述第二关键字块进一步通过关键字元组标识关键字、关键字所对应数据的存储位置。

4、如权利要求3所述的数据存储方法，其特征在于：

通过偏移地址和长度标识所述关键字块或/和数据块的存储位置。

5、如权利要求3所述的数据存储方法，其特征在于：

通过偏移地址和长度标识所述关键字所对应数据的存储位置。

6、一种数据检索方法，其特征在于，包括步骤：

61) 查找索引，根据索引所标识的第一关键字块及其对应数据块的存储位置，直接获取存储的与被检索关键字首字符相同的第一关键字块及其对应数据块，所述索引由提取的关键字的字符组成，所述第一关键字块由首字符相同的关键字组成，所述对应数据块根据所述第一关键字块进行存储；

62) 查找第一关键字块中含有的所述被检索关键字所对应数据的存储位置，依据该位置从所述数据块中获取数据。

7、如权利要求6所述的数据检索方法，其特征在于，所述索引通过标识

元组标识第一关键字块和对应数据块的存储位置, 所述步骤 61) 包括:

611) 提取被检索关键字的首字符;

612) 遍历标识元组, 将被检索关键字的首字符与标识元组中标识的字符进行匹配, 若成功, 获取该标识元组中第一关键字块和对应数据块的存储位置;

613) 依据所述第一关键字块和对应数据块的存储位置, 获取第一关键字块和数据块。

8、如权利要求 7 所述的数据检索方法, 其特征在于:

所述第一关键字块中前两位字符相同的关键词组成第二关键字块;

所述步骤 61) 中进一步对索引进行查找, 根据索引进一步所标识的第二关键字块的存储位置获取与被检索关键字前两位字符相同的第二关键字块;

以及步骤 62) 中, 被检索关键字所对应数据的存储位置的获取具体为, 在第二关键字块中遍历关键字元组, 查找所述被检索关键字, 并获取关键字元组中该被检索关键字所对应数据的存储位置。

9、如权利要求 8 所述的数据检索方法, 其特征在于, 所述索引通过标识元组标识第二关键字块的存储位置, 所述进一步对索引进行查找包括步骤:

91) 提取被检索关键字的第二位字符;

92) 遍历标识元组, 将被检索关键字的前两位字符与标识元组中标识的字符进行匹配, 若成功, 获取该标识元组中第二关键字块的存储位置。

10、如权利要求 9 所述的数据检索方法, 其特征在于:

通过偏移地址和长度标识所述关键字块或/和数据块的存储位置。

11、如权利要求 9 所述的数据检索方法, 其特征在于:

通过偏移地址和长度标识所述关键字所对应数据的存储位置。

12、一种数据存储系统, 包括加密校验信息单元、词典信息单元, 其特征在于, 还包括:

索引单元, 用于提取关键字的字符, 存储由所述关键字的字符组成的标识元组的集合, 所述标识元组标识第一关键字块及其对应数据的存储位置;

第一关键字块集合单元, 用于存储由首字符相同的关键词构成的第一关键字块的集合;

数据块集合单元, 用于存储与所述第一关键字块集合单元存储的第一关键

字块所对应的数据块的集合。

13、如权利要求 12 所述的数据存储系统，其特征在于，所述索引单元存储的标识元组通过第一关键字块标识元组标识：

所述第一关键字块的首字符；

所述第一关键字块的存储位置；

所述第一关键字块对应的数据块的存储位置。

14、如权利要求 13 所述的数据存储系统，其特征在于：

所述第一关键字块集合单元存储的第一关键字块进一步将前两位字符相同的关键字组成第二关键字块；

所述索引单元存储的标识元组进一步通过第二关键字块标识元组标识第二关键字块的前两位字符组合、第二关键字块的存储位置；

所述第二关键字块进一步通过关键字元组标识关键字、关键字所对应数据的存储位置。

15、如权利要求 14 所述的数据存储系统，其特征在于：

所述索引单元通过偏移地址和长度标识所述关键字块或/和数据块的存储位置。

16、如权利要求 14 所述的数据存储系统，其特征在于：

所述第一关键字块集合单元通过偏移地址和长度标识所述关键字所对应数据的存储位置。

## 一种数据存储/检索方法及系统

## 技术领域

本发明涉及数据处理技术，尤其涉及一种词典类数据的存储/检索方法及系统。

## 背景技术

随着无线通讯技术的发展，人们希望在手机等无线设备中存储越来越多的有用信息，例如词典类的数据。通常情况下，词典类数据具有如下特征：数据量庞大，每本词典中的关键字唯一，关键字之间可按照一定顺序排列。为了满足不同用户的需求，通常会提供给用户多本不同类型的词典，例如：汉英词典，英汉词典，专业词典等。

现有词典类数据的存储方法，请参阅图1现有技术词典存储示意图，其存储内容包括：加密校验信息、字典信息、关键字、数据，其中关键字包含与关键字对应的每个解释数据的偏移量。存储的方法是首先将词典中所有关键字以及与关键字对应的每个解释数据的偏移量提取出来，再将整个解释数据压缩后作为一个文件保存。

根据现有存储方法，无线设备在进行数据检索时，在所有关键字中找到欲检索的关键字，获取其解释数据的偏移量，再解压缩解释数据所在文件，通过偏移量找到当前关键字的解释数据。请参阅图2现有技术的检索流程图，包括步骤：

S21、遍历每本目标词典，若遍历结束，则结束检索，若遍历未结束，进入步骤 S22；

S22、根据加密校验信息，校验当前词典有效性，若有效，进入步骤 S23，若无效，进入步骤 S21；

S23、提取欲检索的关键字；

S24、遍历词典中所有关键字，看是否找到匹配关键字，若找到进入步骤 S25，若没有找到进入步骤 S21；

S25、根据找到的匹配关键字，获取关键字含有的与关键字对应的每个解释数据的偏移量；

S26、解压缩解释数据所在文件，通过偏移量找到当前关键字的解释数据，取出本词典数据后，进入步骤 S21 进行下一本目标词典的检索。

可以看出，根据现有存储方法，需要在所有关键字中进行查找并且获取解释数据，相当于对整个解释数据文件进行操作，致使手机检索速度慢。

另外，为了提高检索速度，只能缩小词典的检索范围，例如只选取词典中的部分数据存储于无线设备中，或者减少词典的数目。这两种选择，虽然能加快无线设备的处理检索速度，但由于词典的检索范围缩小，不能满足用户日益增长的信息检索需求。

### 发明内容

本发明要解决的技术问题是提供一种数据存储/检索方法及系统，该方法及系统能将词典类数据进行有效的组合，使无线设备提高数据检索速度，满足用户对信息的检索需求。

为解决上述技术问题，本发明的目的是通过以下技术方案实现：

本发明提供一种数据存储方法，包括：提取关键字的字符，将所述提取的关键字的字符作为索引进行存储，所述索引标识第一关键字块及其对应数据的存储位置；存储由首字符相同的关键字组成的第一关键字块；分块存储所述第一关键字块对应的数据。

进一步的，所述索引通过第一关键字块标识元组标识所述第一关键字块的首字符、第一关键字块的存储位置、第一关键字块对应的数据块的存储位置。

进一步的，存储由第一关键字块中前两位字符相同的关键字组成的第二关键字块；所述索引进一步通过第二关键字块标识元组标识第二关键字块的前两位字符组合、第二关键字块的存储位置；所述第二关键字块进一步通过关键字元组标识关键字、关键字所对应数据的存储位置。

通过偏移地址和长度标识所述关键字块或/和数据块的存储位置。

通过偏移地址和长度标识所述关键字所对应数据的存储位置。

本发明提供一种数据检索方法，包括步骤：61) 查找索引，根据索引所标识的第一关键字块及其对应数据块的存储位置，直接获取存储的与被检索关键字首字符相同的第一关键字块及其对应数据块，所述索引由提取的关键字的字符组成，所述第一关键字块由首字符相同的关键字组成，所述对应数据块根据所述第一关键字块进行存储；62) 查找第一关键字块中含有的所述被检索关键字所对应数据的存储位置，依据该位置从所述数据块中获取数据。

进一步的，所述索引通过标识元组标识第一关键字块和对应数据块的存储位置，所述步骤 61) 包括：611) 提取被检索关键字的首字符；612) 遍历标识元组，将被检索关键字的首字符与标识元组中标识的字符进行匹配，若成功，

获取该标识元组中第一关键字块和对应数据块的存储位置；613) 依据所述第一关键字块和对应数据块的存储位置，获取第一关键字块和数据块。

所述第一关键字块中前两位字符相同的关键字组成第二关键字块；所述步骤61) 中进一步对索引进行查找，根据索引进一步所标识的第二关键字块的存储位置获取与被检索关键字前两位字符相同的第二关键字块；以及步骤62) 中，被检索关键字所对应数据的存储位置的获取具体为，在第二关键字块中遍历关键字元组，查找所述被检索关键字，并获取关键字元组中该被检索关键字所对应数据的存储位置。

所述索引通过标识元组标识第二关键字块的存储位置，所述进一步对索引进行查找包括步骤：91) 提取被检索关键字的第二位字符；92) 遍历标识元组，将被检索关键字的前两位字符与标识元组中标识的字符进行匹配，若成功，获取该标识元组中第二关键字块的存储位置。

通过偏移地址和长度标识所述关键字块或/和数据块的存储位置。

通过偏移地址和长度标识所述关键字所对应数据的存储位置。

本发明提供一种数据存储系统，包括加密校验信息单元、词典信息单元，还包括：索引单元，用于提取关键字的字符，存储由所述关键字的字符组成的标识元组的集合，所述标识元组标识第一关键字块及其对应数据的存储位置；第一关键字块集合单元，用于存储由首字符相同的关键字构成的第一关键字块的集合；数据块集合单元，用于存储与所述第一关键字块集合单元存储的第一关键字块所对应的数据块的集合。

进一步的，所述索引单元存储的标识元组通过第一关键字块标识元组标识所述第一关键字块的首字符、第一关键字块的存储位置、第一关键字块对应的数据块的存储位置。

所述第一关键字块集合单元存储的第一关键字块进一步将前两位字符相同的关键字组成第二关键字块；所述索引单元存储的标识元组进一步通过第二关键字块标识元组标识第二关键字块的前两位字符组合、第二关键字块的存储位置；所述第二关键字块进一步通过关键字元组标识关键字、关键字所对应数据的存储位置。

所述索引单元通过偏移地址和长度标识所述关键字块或/和数据块的存储位置。

所述第一关键字块集合单元通过偏移地址和长度标识所述关键字所对应数据的存储位置。

以上技术方案可以看出，本发明具有以下有益效果：

首先，根据现有存储方法，需要在所有关键字中进行查找并且获取解释数据，相当于对整个数据文件进行操作，而本发明以标识元组为索引，将与第一

关键字块对应的数据进行分块压缩并存储,在检索时,通过获取作为索引的标识元组,取出第一关键字块和对应的数据块,并通过获取关键字块中的关键字元组,从第一关键字块对应的数据块中取出数据。也就是说,不用对整个数据文件进行操作,只需解压缩关键字对应数据块,并从中获取数据,从而提高了无线设备的数据检索速度;

进一步的,根据本发明,由于无线设备的数据检索速度提高,所以可增大词典数据检索的范围,增加检索的词典数目,从而满足用户不断增长的信息检索需求。

#### 附图说明

图1是现有技术的词典存储示意图;

图2是现有技术的检索流程图;

图3是本发明的词典存储示意图;

图4是本发明的索引存储示意图;

图5是本发明的第二关键字块存储示意图;

图6是本发明的检索流程图;

图7是本发明的词典存储系统结构图。

#### 具体实施方式

本发明提供的一种数据存储/检索方法及系统,可以使无线设备检索词典类数据时加快检索速度,增大词典数据检索的范围,满足用户对信息的检索需求。

本发明是以每本词典作为存储单位进行存储,一个目标词典对应一个文件。为了便于对本发明进一步理解,下面结合附图对本发明进行详细描述。

本发明数据存储方法的核心思想是:提取关键字的字符作为索引;将首字符相同的关键字组成第一关键字块;将所述第一关键字块对应的数据进行存储。

请参阅图3,是本发明的词典存储示意图。

每本词典的存储数据包括五部分:加密校验信息、词典信息、索引、第一关键字块、数据块。第一关键字块包括第一关键字块1~第一关键字块n,数据块包括数据块1~数据块n。



以下对这五部分进行详细介绍:

### 1、加密校验信息

记录加密方法及验证信息,用于验证每本词典的完整性。如果校验出当前词典无效,说明此词典不完整,在检索过程中将进入下一本目标词典的检索;如果有效,则可以进入检索。

### 2、词典信息

用于记录词典的版本信息、词典类型、词典中最大解释数据块的长度等。通过此信息用户可以更好了解词典概况,决定是否查询此词典。

### 3、索引

存储所有第一关键字块标识元组和第二关键字块标识元组,其存储方法请参阅图4本发明的索引存储示意图。

索引的存储顺序是第一关键字块1标识元组,第一关键字块1中的所有第二关键字块的标识元组,第一关键字块2标识元组,第一关键字块2中的所有第二关键字块的标识元组,以此类推。

可以理解的是,第一关键字块标识元组之间和第一关键字块中的所有第二关键字块的标识元组之间是按字符顺序进行排列,但并不局限于此。

如图4所示,有两个第一关键字块标识元组,第一关键字块标识元组有*i*个第二关键字块标识元组,第二关键字块标识元组有*j*个第二关键字块标识元组,则存储为:第一关键字块1标识元组,第一关键字块1中的第二关键字块1标识元组、第二关键字块2标识元组、第二关键字块*i*标识元组;第一关键字块2标识元组,第一关键字块2中的第二关键字块1标识元组、第二关键字块2标识元组、第二关键字块*j*标识元组。

需要说明的是,索引的存储顺序并不局限于上述方法,也可以是第一关键字块1标识元组,第一关键字块2标识元组,第一关键字块1中的所有第二关键字块的标识元组,第一关键字块2中的所有第二关键字块的标识元组,以此类推。即在图4中,也可存储为第一关键字块1标识元组,第一关键字块2标识元组;第一关键字块1中的第二关键字块1标识元组、第二关键字块2标识元组、第二关键字块*i*标识元组,第一关键字块2中的第二关键字块1标识元组、第二关键字块2标识元组、第二关键字块*j*标识元组。

第一关键字块标识元组的格式为：(a, P, L, Pc, Lc)

a: 标识每个第一关键字块的首字符;

P: 表示以a为首字符的第一关键字块在整本词典中的偏移, 其偏移是以词典首地址为基准的;

L: 表示以a为首字符的第一关键字块的长度;

Pc: 表示以a为首字符的第一关键字块所对应的数据块在整本词典中的偏移, 其偏移是以词典首地址为基准的;

Lc: 表示以a为首字符的第一关键字块所对应的数据块的长度。

第二关键字块标识元组的格式为：(ab, Ps, Ls)

ab: 标识每个第二关键字块的前两位字符;

Ps: 表示以ab为前两位字符的第二关键字块在以a为首字符的第一关键字块中的偏移, 其偏移是以第一关键字块的地址为基准, 解压缩后的偏移;

Ls: 表示以ab为前两位字符的第二关键字块的长度。

如图4所示, 第一关键字块1标识元组含有(a, P1, L1, Pc1, Lc1), 第二关键字块1标识元组含有(aa, Ps1, Ls1), 第二关键字块2标识元组含有(ab, Ps2, Ls2), 第二关键字块i标识元组含有(az, Psi, Lsi), 第一关键字块2标识元组含有(b, P2, L2, Pc2, Lc2), 第二关键字块1标识元组含有(ba, Ps1, Ls1), 第二关键字块2标识元组含有(bb, Ps2, Ls2), 第二关键字块j标识元组含有(bz, Psj, Lsj)。

#### 4、第一关键字块

第一关键字块包括第一关键字块1~第一关键字块n。

第一关键字块由第二关键字块组成, 是首字符相同的第二关键字块的集合。在一本词典中, 若有n个不同的首字符, 就有n个第一关键字块。对每个第一关键字块, 进行压缩后存储。

第二关键字块是前两个字符相同的所有关键字的集合, 还包括关键字为一个字符的关键字集合, 第二关键字块存储关键字元组。当前关键字块中有j个关键字, 就有j个关键字元组。

关键字元组的存储格式为：(Str, Pk, Lk)

Str: 表示关键字;

Pk: 表示当前关键字所对应数据在数据块中的偏移, 其偏移是以数据块的首地址为基准, 解压缩后的偏移。

Lk: 表示关键字对应数据的长度。

例如关键字元组 (about, Pk, Lk), 其中about表示关键字, Pk表示关键字about对应的数据在所属第一关键字块对应的数据块中的偏移, 其偏移是以数据块的首地址为基准, 解压缩后的偏移, Lk表示about对应数据的长度。

请参阅图5, 是本发明的第二关键字块存储示意图。

图5中的41表示单字符a的第二关键字块, a表示关键字, Pk1表示a对应数据在所属第一关键字块对应的数据块中的偏移, 其偏移是以数据块的首地址为基准, 解压缩后的偏移, Lk1表示关键字a对应数据的长度;

42表示以ab开头的第二关键字块, 如abacus表示关键字, Pk21表示abacus对应数据在所属第一关键字块对应的数据块中的偏移, 其偏移是以数据块的首地址为基准, 解压缩后的偏移, Lk21表示abacus对应数据的长度;

43表示以ac开头的第二关键字块; 如academic表示关键字, Pk31表示academic对应数据在所属第一关键字块对应的数据块中的偏移, 其偏移是以数据块的首地址为基准, 解压缩后的偏移, Lk31表示academic对应数据的长度。

其他情况以此类推。

一般情况下, 第一关键字块不存储关键字, 当关键字是一个字符时, 例如单词a, 则存储a的关键字元组在第一关键字块中, 与在第二关键字块中是并列的地位。

需要说明的是, 由于每个第一关键字块所包括的第二关键字块的数量不同, 并且每个第二关键字块的长度各不相同, 其长度取决于每个关键字的长度和包括的关键字的数量, 因此每个第一关键字块不一定占据相同的空间。

可以看出, 此处提到的关键字元组和前面提到的索引中的标识元组的区别是: 关键字元组存储的是各个具体的关键字, 标识元组只是用于标识关键字块中存储的关键字的首位及前两位字符, 还有相应的偏移。

## 5、数据块

数据块包括数据块 1 ~ 数据块 N。

每个数据块存储每个第一关键字块中所有第二关键字块的关键字所对应

的数据。有多少个第一关键字块，就有多少个数据块。每个数据块压缩后进行存储。

本发明数据检索方法的核心思想包括步骤：1) 查找索引，获取与被检索关键字首字符相同的第一关键字块和对应数据块；2) 查找第一关键字块中所述被检索关键字所对应数据的存储位置，依据该位置从所述数据块中获取数据。

请参阅图6，是本发明的检索流程图。

在检索开始前，先获取用户欲检索的关键字Str，获取方式包括通过屏幕取词或者用户输入等，然后依据预先设置的目标词典ID，找到目标词典所存储的文件。一般一个文件对应一个目标词典。

此时进入检索流程，如图6所示，包括步骤：

S51、遍历每本目标词典，若遍历结束，则结束检索，若遍历未结束，进入步骤S52，遍历词典的数目是根据用户的设置，或者是程序预先的设置，例如共有10本词典，用户需要获取关键字在其中3本词典的解释，那么就检索那3本词典；

S52、根据加密校验信息，首先验证当前词典文件是否完整有效，若有效，进入步骤S53，若无效，进入步骤S51进行下一本目标词典的检索；

S53、提取关键字Str的第一个字符S1；

S54、遍历索引中第一关键字块标识元组，将Str的第一个字符S1与第一关键字块标识元组中的首字符进行匹配，若匹配成功，进入步骤S55，若匹配失败，进入步骤S51进行下一本目标词典的检索；

S55、获取第一关键字块标识元组X，得到相应第一关键字块的首字符S1、第一关键字块在整本词典中的偏移P、第一关键字块的长度L、第一关键字块所对应的数据块在整本词典中的偏移Pc、第一关键字块所对应的数据块的长度Lc；

S56、依据第一关键字块标识元组X中的P和L，取出Str所属第一关键字块，进行解压缩，依据第一关键字块标识元组X中的Pc和Lc，取出第一关键字块对应的数据块，进行解压缩；

S57、提取关键字Str的第二个字符S2，如果关键字只有一个字符S1时，

进入步骤 S61;

S58、遍历索引中上述第一关键字块标识元组后的第二关键字块标识元组,将关键字Str前两位字符S1, S2与第二关键字块标识元组中的字符进行匹配,若匹配成功,进入步骤S59,若匹配失败,进入步骤S51进行下一本目标词典的检索;

S59、获取第二关键字块标识元组Y,得到相应第二关键字块的前两位字符S1S2、第二关键字块在以S1为首字符的第一关键字块中的偏移Ps及第二关键字块的长度Ls;

S60、依据第二关键字块标识元组Y中的Ps和Ls,在已获取的第一关键字块中获取Str所属第二关键字块i;

S61、在第二关键字块i中,包括关键字只有一个字符S1时,遍历所有关键字元组,与str进行匹配,匹配成功,进入步骤S62,若匹配失败,进入步骤S51进行下一本目标词典的检索;

S62、获取关键字Str对应的关键字元组Z,得到相应关键字元组关键字Str,当前关键字所对应数据在数据块中的偏移Pk、关键字对应数据的长度Lk;

S63、在已获取的解压缩后的第一关键字块对应的数据块中,依据获取的关键字元组Z中的Pk和Lk,取出关键字str对应的数据,取出本词典数据后,进入步骤S51进行下一本目标词典的检索。

当所有目标词典检索完毕后,将检索到的每本目标词典中的解释数据反馈给用户。

请参阅图7,是本发明的词典存储系统结构图。

数据存储系统 600 包括: 加密校验信息单元 601、词典信息单元 602、索引单元 603、第一关键字块集合单元 604、数据块集合单元 605。

加密校验信息单元 601 记录加密方法及验证信息,用于验证每本词典的完整性。如果校验出当前词典无效,说明此词典不完整,将进入下一本目标词典的检索,如果有效,则可以进入检索。

词典信息单元 602 用于记录词典的版本信息、词典类型、词典中最大解释数据块的长度等。通过此单元信息用户可以更好了解词典概况,决定是否查询此词典。

索引单元 603 存储所有第一关键字块标识元组和第二关键字块标识元组。第一关键字块标识元组含有：标识每个第一关键字块的首字符  $a$ ；以  $a$  为首字符的第一关键字块在整本词典中的偏移  $P$ ，其偏移是以词典首地址为基准的；以  $a$  为首字符的第一关键字块的长度  $L$ ；以  $a$  为首字符的第一关键字块所对应的数据块在整本词典中的偏移  $P_c$ ，其偏移是以词典首地址为基准的；以  $a$  为首字符的第一关键字块所对应的数据块的长度  $L_c$ 。第二关键字块标识元组含有：标识每个第二关键字块的前两位字符  $ab$ ；以  $ab$  为前两位字符的第二关键字块在以  $a$  为首字符的第一关键字块中的偏移  $P_s$ ，其偏移是以第一关键字块的地址为基准，解压缩后的偏移；以  $ab$  为前两位字符的第二关键字块的长度  $L_s$ 。根据存储的第一关键字块标识元组，可以获取第一关键字块集合单元 604 存储的第一关键字块和数据块集合单元 605 存储的与第一关键字块对应的数据块，根据存储的第二关键字块标识元组，可以在第一关键字块集合单元 604 存储的第一关键字块中获取第二关键字块。

第一关键字块集合单元 604 存储第一关键字块，包括第一关键字块 1 ~ 第一关键字块  $n$ 。第一关键字块由第二关键字块组成，是首字符相同的第二关键字块的集合。第二关键字块是前两个字符相同的所有关键字的集合，还包括关键字为一个字符的关键字集合，第二关键字块存储关键字元组。第二关键字块存储的关键字元组含有：关键字  $Str$ ；表示当前关键字在所属第一关键字块所对应的数据块中的偏移  $P_k$ ，其偏移是以数据块的首地址为基准，解压缩后的偏移；关键字所对应数据的长度  $L_k$ 。根据第二关键字块存储的关键字元组，可以从数据块集合单元 605 中存储的与第一关键字块对应的数据块中取出数据。

数据块集合单元 605 存储与第一关键字块集合单元 604 对应的数据，包括数据块 1 ~ 数据块  $N$ 。具体来说，所述数据是与第一关键字块集合单元 604 中第一关键字块含有的第二关键字块中所有关键字对应。

综上所述，本发明将词典中第一关键字块标识元组和第二关键字块标识元组作为索引，按第一关键字块、第一关键字块对应的数据块进行组织，并以第一关键字块为单位，对第一关键字块中的第二关键字块所有的关键字的数据解释，进行分块组织并压缩存储，因此在进行检索时，只需解压缩一小块数据获取所需信息，从而提高了检索速度，增大了词典数据检索的范围，满足了用户

的检索需求。

以上对本发明所提供的一种数据存储/检索方法及系统进行了详细介绍，本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

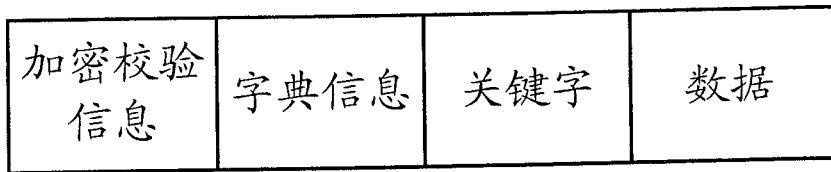


图 1

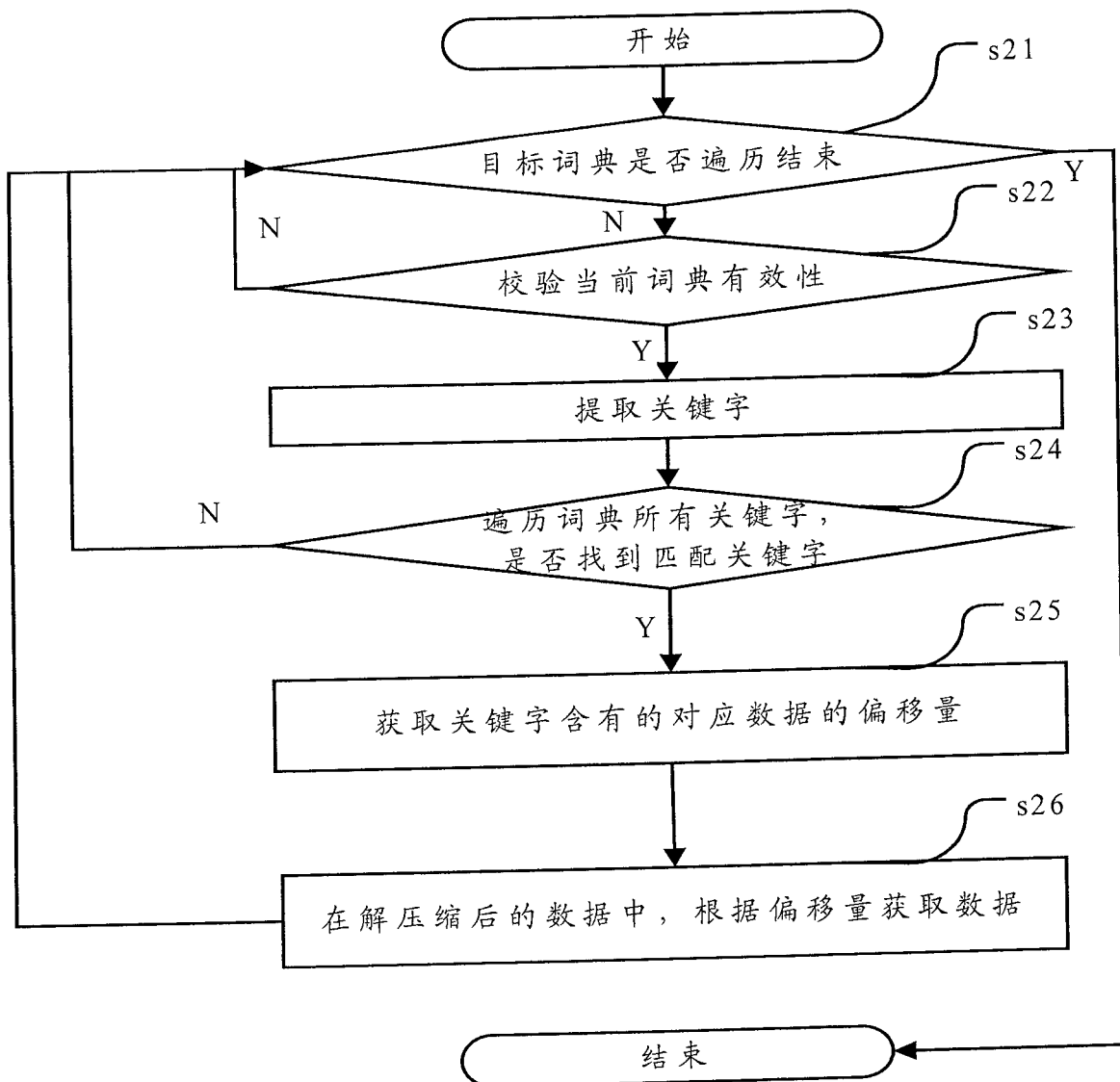


图 2



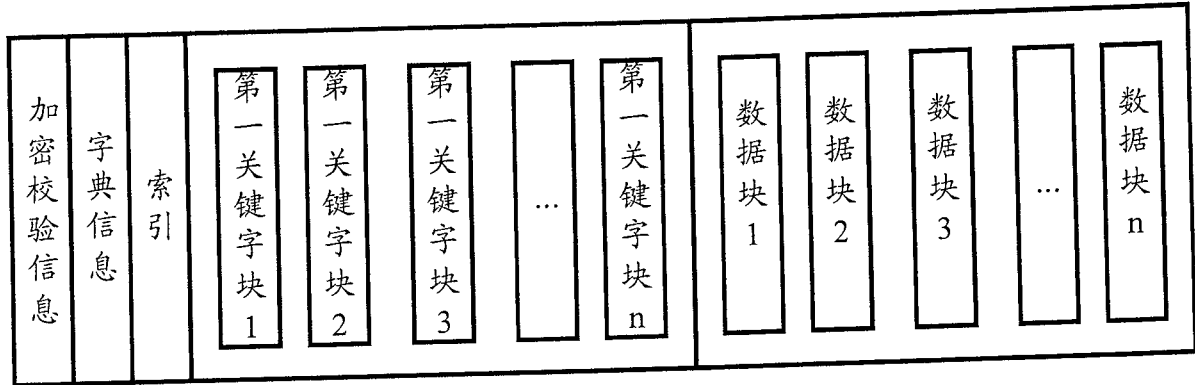


图 3

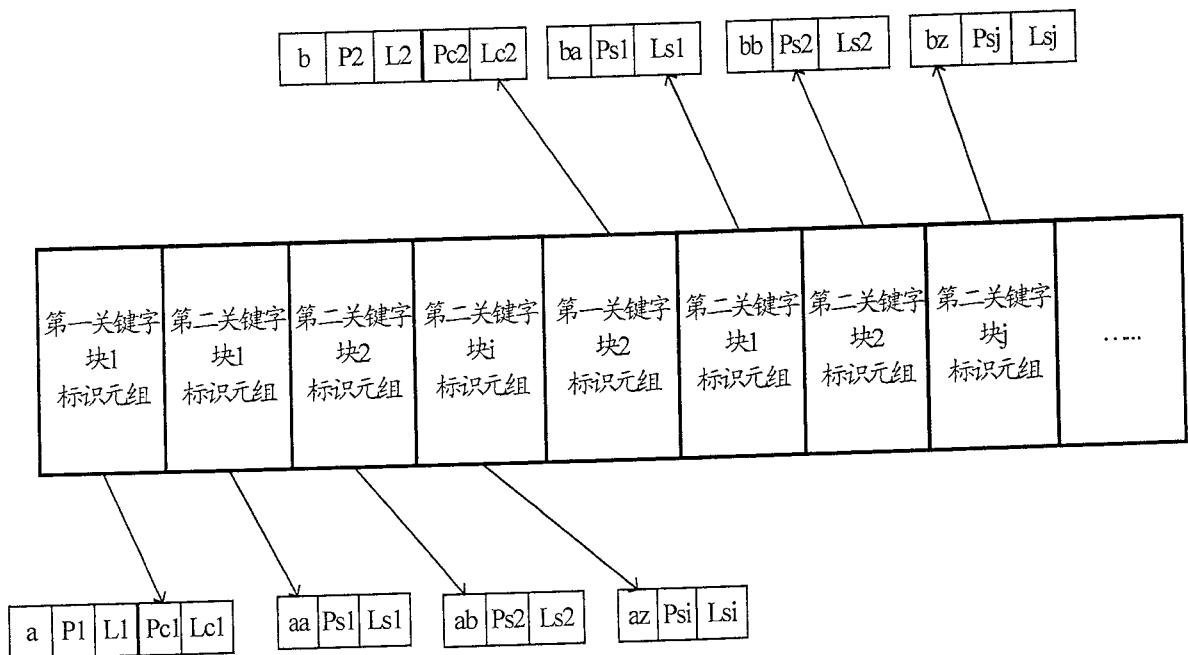


图 4

a	Pk1	Lk1	41
abacus	Pk21	Lk21	
abandon	Pk22	Lk22	42
... ..			
abyss	Pk2k	Lk2k	
academic	Pk31	Lk31	43
... ..			
acute	Pk3k	Lk3k	

图 5

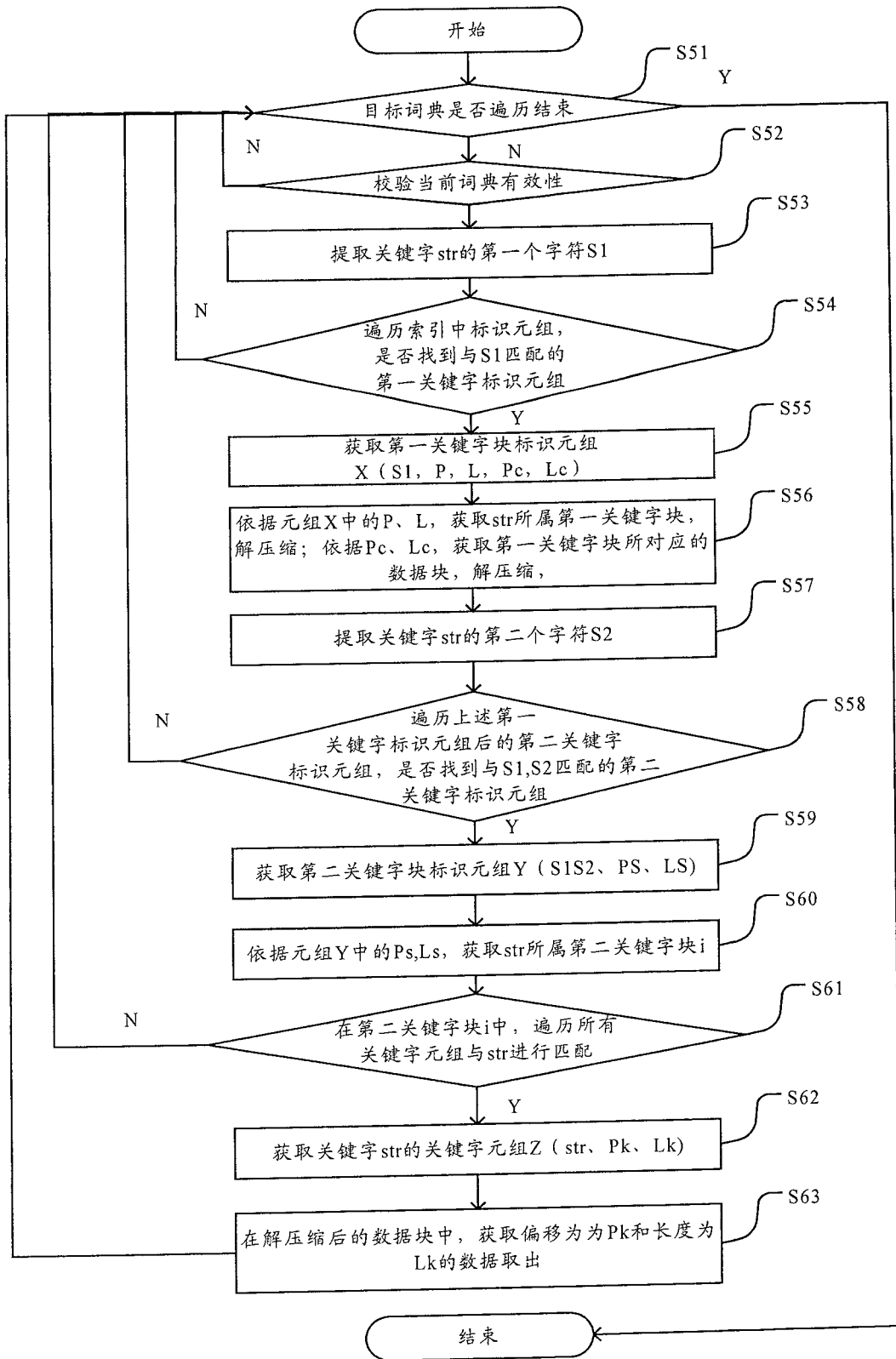


图 6

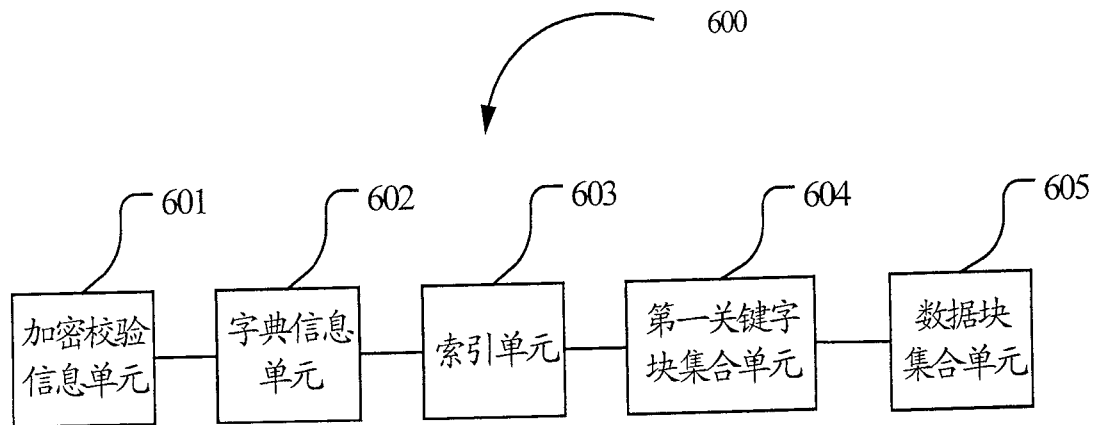


图 7