

# 一种基于编码压缩的数据广播关键字索引方法

张 健, 孙未未

(复旦大学计算机科学技术学院, 上海 201203)

**摘 要:** 无线环境的特殊性导致传统的关键字检索方法不能很好地用于周期数据广播之中。倒排表是全文检索中广泛使用的一种索引技术, 但倒排表索引和基于哈希的数据索引无法解决索引结构过大的问题。为此, 在周期数据广播环境下, 提出一种新型的关键字索引结构, 对倒排表进行编码压缩, 缩减索引结构来减少访问时间和调谐时间。同时, 与编码压缩索引相结合, 设计一种周期数据广播下的文档调度方法。在真实数据集上进行的实验结果表明, 该方法可缩减索引结构的规模, 降低访问延迟和能耗。

**关键词:** 无线环境; 数据广播; 关键字检索; 索引; 编码压缩; 倒排表

**中文引用格式:** 张 健, 孙未未. 一种基于编码压缩的数据广播关键字索引方法[J]. 计算机工程, 2015, 41(1): 75-81.

**英文引用格式:** Zhang Jian, Sun Weiwei. A Keyword Index Method for Data Broadcast Based on Coding Compression[J]. Computer Engineering, 2015, 41(1): 75-81.

## A Keyword Index Method for Data Broadcast Based on Coding Compression

ZHANG Jian, SUN Weiwei

(School of Computer Science, Fudan University, Shanghai 201203, China)

**【Abstract】** The traditional keyword index methods are unable to be properly applied in the wireless data broadcast. Inverted list is an index technique widely used in the keyword search. Inverted list index and hash-based stream index can not be able to deal with the problem of a too large index structure. This paper proposes a new type of keyword index structure, and it manages to make the index structure smaller and shorten the tuning time by a way of coding compression. At the same time, it proposes a document scheduling method in the environment of periodic data broadcast. The experimental results demonstrate that the index structure is latency and energy-efficient, and outperforms inverted list index and hash-based stream index.

**【Key words】** wireless environment; data broadcast; key word search; index; coding compression; inverted list

DOI: 10.3969/j.issn.1000-3428.2015.01.014

### 1 概述

随着无线设备的快速推广和第 4 代无线网络(4G)的迅速崛起, 无线通信由于其灵活方便的特性而变得越来越重要。无线通信中的有限的带宽和无线设备的电池续航能力成为了移动计算中较为主要的关注点, 这使无线数据广播会变成移动通信领域一个比较重要的数据传输技术。一般来说, 按照调度模式的不同, 无线数据广播可分成 2 种类型: On-demand 模式和周期广播模式<sup>[1]</sup>。对于前者, 用户通过上行信道将请求发送给服务器端, 服务器根据所有用户的情况安排调度数据到信道当中; 在后者中, 服务器端将存储的数据按照某种调度方式在广播信

道上循环广播, 用户端只需在广播信道上侦听, 一旦发现自己感兴趣的数据则下载到本地。本文的索引结构面向的是周期数据广播。周期数据广播技术具有如下的优点: (1) 周期数据广播具有很高的带宽利用率, 它充分利用了广播信道的带宽作为下行信道, 而且对上行信道几乎没有影响; (2) 周期数据广播具有良好的扩展性, 能够支持一个区域内任意数量的用户同时访问数据; (3) 周期数据广播对用户的隐私提供很好的保护<sup>[2]</sup>。

周期数据广播技术的特点能够使它很好地利用到无线环境的数据通信中, 而且数据广播特别适用于公共信息的发布, 如新闻广播和地理交通信息, 因为面向的用户群体非常大。基于关键字的查询作为

**基金项目:** 国家自然科学基金资助项目(61073001)。

**作者简介:** 张 健(1988 - ), 男, 硕士研究生, 主研方向: 移动数据管理; 孙未未, 副教授。

**收稿日期:** 2014-02-25    **修回日期:** 2014-03-23    **E-mail:** zhang\_jian@fudan.edu.cn

一种常用的查询方式成为了数据广播的一个重要的研究问题。

周期广播通常会使用一定的索引方法,把索引和数据一同插入到广播的数据流中。通过索引信息,用户被告知自己需求的数据包的到达时间,跳过不必要的下载。移动设备可在无关数据的广播时隙启用休眠模式,节省能耗。无线数据广播中有 2 个主要的性能指标:访问时间(Access Time)和调谐时间(Tuning Time),分别定义为从用户提交查询到最终获得查询结果的时间间隔和用户保持监听状态的时间。

关于文本数据中关键字检索的问题,目前已有大批的索引结构和查询方法<sup>[3-8]</sup>。另一方面,在文本检索中,一个关键字往往对应多个文档,可以将文档调度问题等同于数据广播中的多数据项广播调度问题(如文献[9-11])。但已有的数据调度算法都存在一定的局限性。为此,本文提出无线数据广播环境下基于编码压缩的关键字查找索引方法,并与新型关键字索引相结合,对周期数据广播中的文档进行调度。

## 2 相关研究

目前关于文本数据的索引结构和查询方法较多,例如,文献[3]提出了能增量更新的查询索引;文献[4]讨论了关键字检索中的索引压缩方法;文献[5]研究了文本搜索引擎中倒排表的应用;文献[6]提出分布式系统的文本检索优化方法,但这些方法都是基于磁盘访问数据的,而在无线广播环境中,文档是根据时间线性地存储在广播信道中的,传统方法不能够直接地应用于无线环境中。文献[7]针对无线环境提出了 B+-tree-倒排表的两层索引结构和 $(1, \alpha(1, \beta))$ 的分布策略。文献[8]把关键字索引按哈希值映射到数据流中,并根据数据广播的特点提出了优化后效率较高的 Merged-Hash 索引方法。本文提出了一种新型的索引方法,它对倒排表进行编码压缩,大大减小了索引结构,提高了系统的性能。

另一方面,可以把文档调度问题等同于数据广播中的多数据项广播调度问题,例如,文献[9]提出基于数据项的访问频率进行调度以减少访问延迟;文献[10]针对多数据项广播结合访问频率和请求需要来生成调度序列;文献[11]就多信道广播环境下提出了数据项的广播调度方法。已有的数据调度算法都存在一定的局限性:服务器必须事先知道请求的分布情况;限制了系统在实时环境下调度算法的实用性;文献[12]提到了数据广播中通过比较 XML 文档的相似性来对文档进行调度;文献[13]通过合

并亲密度高的文档以减少冗余信息。不过半结构化的 XML 文档与非结构化的纯文本数据有所区别,不能直接使用到纯文本数据的调度中。为此,本文提出一种基于编码压缩的数据广播关键字索引方法。

## 3 编码压缩索引

### 3.1 数据广播中的关键字查询

为简化模型,本文考虑只具有一个基站和用户共同使用一个广播传输信道的场景。在广播数据过程中,广播的内容在相当长的一段时间不会发生更新。基站会在每一个广播周期广播若干个文档。每个文档只在该广播周期中出现一次,文档集合记作  $D = \{doc_1, doc_2, \dots, doc_n\}$ 。每个文档  $doc_i$  都会通过若干个数据包发送到广播信道中。在此,数据包是广播信道中最小的逻辑单元,每个数据包都具有相同的大小。为了方便,表 1 列出了下文中用到的符号。

表 1 符号说明

符号	描述
$D$	文档集合 $\{doc_1, doc_2, \dots, doc_n\}$
$n$	文档集合数目
$K$	关键字集合 $\{k_1, k_2, \dots, k_v\}$
$v$	文档集合中所有关键字的数目
$DK_i$	$doc_i$ 包含的关键字集合
$KD_j$	包含关键字 $k_j$ 的文档集合
$QS$	查询关键字集合
$R_{QS}$	同时满足 $QS$ 关键字集合的文档集合

在文本检索的应用中,每个关键字都可能对应数据流中的多个文档。或者说,一个文档会包含多个关键字。为了提供对这种关键字和文档之间的多对多关系的支持,倒排表就被提出来应用于文档检索系统中。倒排表的每个表项都包含一个关键字和包含该关键字的文档的地址。在数据传输过程中,服务器通过流的形式来广播倒排表索引和文档。

通过上述介绍可以直观地看出,每个关键字都是直接和文档的地址相联系起来的。如果文档的数目比较大,而且部分关键字频繁在不同的文档中出现,那么这部分的关键字对应的表项将会有有一个非常长的文档地址列表。这样会导致整个索引结构变得比较大,对整个系统的性能特别是访问时间会带来负面影响。

本文提出了一个改进的方法来克服原来的倒排列表在索引构建方面的缺陷。相比于原来的倒排表索引,该方法采取了另一种方法来表示某个关键字对应的文档地址列表。

### 3.2 二元组

本文把文档 ID 按照某种顺序排列放到一个序

列中, 记作  $S = \{i_1, i_2, \dots, i_n\}$ 。举例假设序列是按照 ID 的大小顺序依次放置的, 即  $i_j = j$ 。相比于倒排索引中关键字直接对应到文档偏移位置, 编码压缩索引中的关键字  $k_i$  是和一组偏移量二元组相联系的, 记作  $F_i = \{ \langle s_{i1}, e_{i1} \rangle, \langle s_{i2}, e_{i2} \rangle, \dots, \langle s_{in}, e_{in} \rangle \}$ 。二元组  $\langle s_{i1}, e_{i1} \rangle$  表示的是在序列  $S$  中下标由  $s_{i1}$  到  $e_{i1}$  的所有文档, 这些文档都包含关键字  $k_i$ 。

可直观发现, 通过调整序列  $S$  中文档的排列顺序可以改变集合  $F_i$  中二元组的数目。面临的问题是怎么样去安排序列  $S$  中文档 ID 的排列顺序能够使集合  $F_i$  中的二元组的数目最少, 因为缩减二元组数目的大小能够有效地缩减索引大小, 从而减少访问时间。

### 3.3 最优序列构造

需求解的问题是如何选择序列  $S$ , 使得  $\sum_{i=1}^v |F_i|$  达到最小。构造一个图  $G(V, E)$ , 图上的每个节点都代表集合  $D$  中的一个文档。也就是说, 对于任何  $doc_j \in D, 1 \leq j \leq n$ , 都有一个对应的  $v_j \in V(G)$ 。对于任意 2 个节点  $v_i$  和  $v_j$  中间都有一条无向边  $e_{ij} = (v_i, v_j) \in E(G)$ , 边的权值  $w(e_{ij})$  是该 2 个节点对应的文档所拥有的共同的关键字的数目。

定义  $\theta(S) = \{ (v_i, v_j) \mid doc_i \text{ 和 } doc_j \text{ 在序列 } S \text{ 中相邻} \}$ 。由于序列  $S$  中没有重复的文档, 故  $|S| = |V|$ , 因此  $\theta(S)$  构成图  $G$  的一条哈密顿回路。定义是关键字  $k_i$  对应文档集合中任意 2 个文档在序列  $S$  中是相邻的次数。例如序列  $S = \{doc_1, doc_2, doc_3, doc_4, doc_5\}$ ,  $KD_1 = \{doc_1, doc_2, doc_5\}$ , 其中在序列  $S$  相邻的文档对有  $(doc_1, doc_2)$  和  $(doc_2, doc_5)$ , 可得  $M_S(KD_1) = 2$ 。

#### 引理 1

$$\sum_{i=1}^v M_S(KD_i) = \sum_{e_{ij} \in \theta(S)} w(e_{ij}) \quad (1)$$

证明: 定义

$$\varphi_S(doc_j, doc_k) = \begin{cases} 1 & doc_j \text{ 与 } doc_k \text{ 在 } S \text{ 中相邻} \\ 0 & \text{其他} \end{cases} \quad (2)$$

$$\sigma_i(doc_j) = \begin{cases} 1 & doc_j \in KD_i \\ 0 & doc_j \notin KD_i \end{cases} \quad (3)$$

$$\tau_j(k_i) = \begin{cases} 1 & k_i \in DK_j \\ 0 & k_i \notin DK_j \end{cases} \quad (4)$$

易知:

$$\sigma_i(doc_j) = \tau_j(k_i)$$

$$w(e_{jk}) = \sum_{i=1}^v \sigma_i(doc_j) \sigma_i(doc_k)$$

$$M_S(KD_i) = \sum_{j=1}^n \tau_j(k_i) \sum_{k=j+1}^n \tau_k(k_i) \varphi_S(doc_j, doc_k)$$

从而可得:

$$\begin{aligned} \sum_{j=1}^n M_S(KD_i) &= \\ \sum_{i=1}^v \sum_{j=1}^n \tau_j(k_i) \sum_{k=j+1}^n \tau_k(k_i) \varphi_S(doc_j, doc_k) &= \\ \sum_{i=1}^v \sum_{j=1}^n \sigma_i(doc_j) \sum_{k=j+1}^n \sigma_i(doc_k) \varphi_S(doc_j, doc_k) &= \\ \sum_{j=1}^n \sum_{k=j+1}^n \varphi_S(doc_j, doc_k) w(e_{jk}) &= \\ \sum_{e_{ij} \in \theta(S)} w(e_{ij}) \end{aligned}$$

#### 引理 2

$$|F_i| = |KD_i| - M_S(KD_i), \quad 1 \leq i \leq t \quad (5)$$

证明: 设  $F_i = \{ \langle s_{i1}, e_{i1} \rangle, \langle s_{i2}, e_{i2} \rangle, \dots, \langle s_{im}, e_{im} \rangle \}$ , 可以计算得到:

$$|KD_i| = \sum_{j=1}^m (e_{ij} - s_{ij} + 1)$$

由于对于任意  $s_j \leq k \leq e_j$  和任意的  $l \neq j$ , 必然有  $doc_k$  不与  $\langle s_l, e_l \rangle$  中表示的任意文档相邻。否则  $\langle s_j, e_j \rangle$  和  $\langle s_l, e_l \rangle$  可以合并为一个二元组。而在  $\langle s_j, e_j \rangle$  中, 相互相邻的文档对数为  $e_j - s_j$ 。证明得:

$$\begin{aligned} |KD_i| - M_S(KD_i) &= \sum_{j=1}^m (e_{ij} - s_{ij} + 1) - \\ &\sum_{j=1}^m (e_{ij} - s_{ij}) = m = |F_i| \end{aligned}$$

#### 定理

$$\sum_{1 \leq i \leq v} |F_i| = \sum_{1 \leq j \leq n} |DK_j| - \sum_{e_{ij} \in \theta(S)} \theta(S) w(e_{ij}) \quad (6)$$

证明: 根据引理 1 和引理 2 可以推出:

$$\begin{aligned} \sum_{1 \leq i \leq v} |F_i| &= \sum_{1 \leq i \leq v} (|KD_i| - M_S(KD_i)) = \\ &\sum_{1 \leq j \leq n} |DK_j| - \sum_{e_{ij} \in \theta(S)} w(e_{ij}) \end{aligned}$$

由于  $\sum_{1 \leq j \leq n} |DK_j|$  是与序列无关且保持不变的, 因此当  $\sum_{e_{ij} \in \theta(S)} w(e_{ij})$  最大时,  $\sum_{1 \leq i \leq t} |F_i|$  达到最小。所以, 构造最优序列  $S$  的问题转化成了求连通图  $G$  上的权值和最大的哈密顿回路, 亦即 TSP 问题。计算的复杂性研究表明, TSP 问题是一个典型的 NP 难问题<sup>[14]</sup>。对于大规模 TSP 问题, 人们目前无法找到完美的求解算法, 所有的智能算法都只能在一定精确范围内求解。本文选取了求解 TSP 问题的开源算法 concorde<sup>[15]</sup> 来解决该问题, 即通过求解等价的 TSP 问题来获得最优的文档序列。

### 3.4 文档调度

假设查询关键字  $k_i$  对应的文档个数为  $N$ , 广播周期长度是  $L$ , 而且每个文档的固定长度记作  $L_D$ , 则满足查询关键字的文档集合的总长度是  $N \times L_D$ 。又假设文档随机分布在信道上,  $doc_j$  和  $doc_{j+1}$  都属于满足关键字  $k_i$  的查询文档, 分布情况如图 1(a) 所示。现把  $doc_j$  和  $doc_{j+1}$  连续放置, 如图 1(b) 所示。按照查询随机进入的条件, 平均的访问时间的差值为:



$$\Delta_{AT} = \frac{gap_j + gap_{j+1}}{2L} \left( L - \frac{gap_j + gap_{j+1}}{2} \right) - \left( \frac{gap_j}{2L} \times \left( L - \frac{gap_j}{2} \right) + \frac{gap_{j+1}}{2L} \left( L - \frac{gap_{j+1}}{2} \right) \right) = -\frac{gap_j gap_{j+1}}{2L} < 0 \quad (7)$$

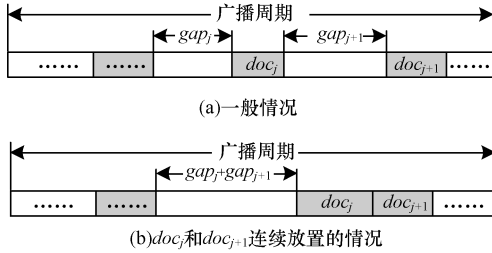


图 1 文档调度

可以看出,把具有相同关键字的文档毗邻放置时,平均访问时间会相比下降。最优的情况是查询关键字  $k_i$  对应的文档被连续广播。此时的平均访问时间为:

$$AT_{avg} = \frac{L - N \times L_D}{L} \left( \frac{1}{2} (L - N \times L_D) + N \times L_D \right) + \frac{N \times L_D}{L} \times L = \frac{L}{2} + N \times L_D - \frac{N^2 \times L_D^2}{2L} \quad (8)$$

在没有进行文档的调度,只是随机地把文档插入到数据流中,可以近似看成文档是均匀分布,此时的  $k_i$  的平均访问时间为:

$$AT'_{avg} = \frac{L - N \times L_D}{L} \left( L - \frac{L - N \times L_D}{2N} \right) + \frac{N \times L_D}{L} \times L = \frac{2N - 1}{2N} L + L_D - \frac{L_D^2}{2L} \quad (9)$$

当  $(N \times L_D)$  较小时,访问时间  $AT'_{avg}$  接近于广播周期  $L$ ,而进行过调度后的  $AT_{avg}$  接近于  $L/2$ ,访问时间会大大减少。普遍来说不能做到对于所有关键字都能把相应的文档聚集到相邻位置,尽可能地把满足同一关键字查询的文档归置到一起有利于减少访问时间,而在解决寻求最优序列使得  $\sum_{1 \leq i \leq t} |F_i|$  最小的过程中,已将带有相同关键字的文档尽可能地放置在一起,从而使访问时间接近  $AT_{avg}$ 。

### 3.5 索引构造

索引构造先构建倒排表,同时对于所有的文档求出它们之间共有的关键字数目,并根据其结果构造图  $G(V, E)$ ,求出最优的哈密顿回路。此时获得哈密顿回路上的节点对应的文档序列。根据文档序列来把原始的倒排表转化为编码压缩的表示。

索引部分按照获得的序列把文档按顺序排放到数据流中,按照  $(1, \alpha(1, \beta))$  的方法把索引树和倒排表的索引插入到数据流中并广播出去。在二元组

$\langle s, e \rangle$  的具体表示方面,本文定义了 2 种指针表示方法,一种是如果  $s = e$ ,即表示单个文档时,用单个指针来表示这个文档,否则就用  $\langle Offset_1, Offset_2 \rangle$  的形式来表示,分别表示文档区间中的第一个和最后一个文档的到达时间。而在这 2 个时间之间所有到达的文档都是符合查询关键字的文档。具体实现如图 2 所示。

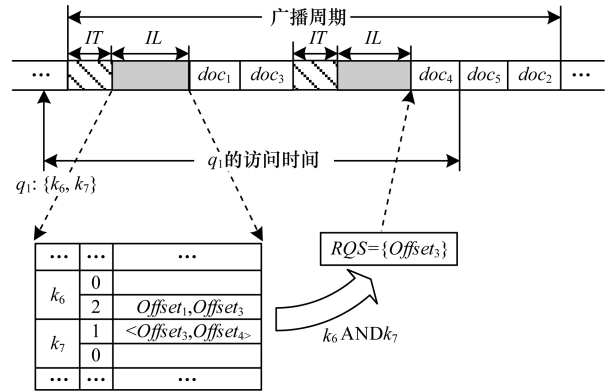


图 2 查询处理过程

### 3.6 用户查询处理

查询过程中用户首先进入信道下载第一个包,如果不是索引树根节点,则等待到下一个索引树开端的索引包到达时下载。然后用户在索引树上搜索自己需要的节点。对于索引树节点,用户会把包含  $QS$  中关键字的孩子节点或对应倒排表表项的指针插入到待下载队列中;而对于倒排表节点,用户直接读出文档到达时间的区间,然后对关键字对应的文档集合取交集。最后用户把文档集合的所有数据包都下载下来。

## 4 实验与结果分析

### 4.1 实验配置

本节通过实验对编码优化倒排表索引(CCKI)、倒排索引(IL)<sup>[7]</sup>和哈希索引(MBHS)<sup>[8]</sup>的性能进行了比较。由于数据广播可广泛地使用于包括新闻和交通信息等方面的公共信息的发布,因此本文实验采用 2 个不同类型的数据集:澳大利亚地图数据集(AU)<sup>[16]</sup>和《洛杉矶时报》的新闻文章(LA)<sup>[17]</sup>。2 个数据集的详细信息如表 2 所示。

表 2 数据集详细信息

数据集	总数	文档平均关键字数目	平均文档大小/Byte
AU	69 745	4.05	256.00
LA	14 478	2 480.50	2 816.82

模拟实验比较的性能包括访问时间(AT)和调谐时间(TT),实验中使用的数据的默认参数值如表 3 所示。

表 3 实验默认参数说明

参数集	平均查询关键字数目	文档数目	每个周期内的查询数目
AU	3	30 000	20 000
LA	3	6 000	20 000

实验模拟进行组织广播索引和数据,并周期性地通过一个信道将数据广播给用户。实验中有多个用户在周期内的任意一个时刻进入信道,他们的查询包括随机生成的多个关键字。在下载完请求的文档后,统计该用户的访问时间和调谐时间。

4.2 性能比较与分析

本文首先通过实验评估对索引进行编码压缩的

效果;然后验证提出的文档调度算法的优化效果;最后分别针对不同文档数目和不同关键字个数,对编码压缩索引和倒排索引、哈希索引进行访问时间和调谐时间的性能比较。

**实验 1** 比较经过编码压缩前后的倒排表大小和原倒排表大小。由图 3 可以看出,经过压缩后,索引大小相比原始倒排表明显降低,而且随着文档数目的增多,压缩效率更高。AU 地图数据集的压缩率在 55% ~ 67% 的范围内,LA 数据集文档含关键字的数目较多且分布相对比较均匀,压缩效果稍差,不过压缩率也在 74% ~ 76% 范围内。

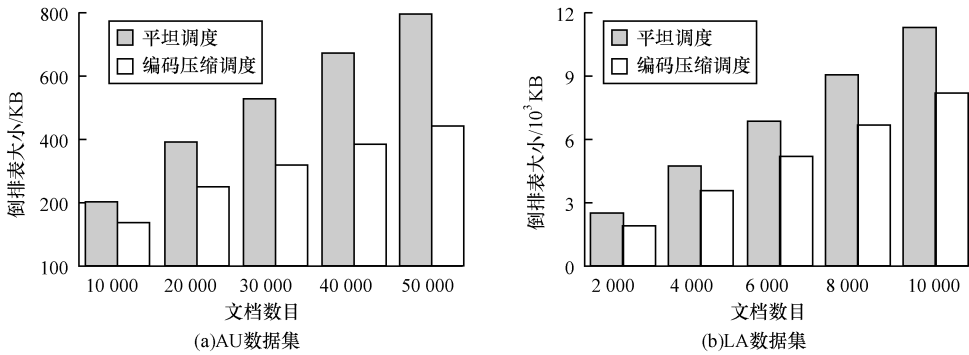


图 3 压缩前后倒排表大小比较

**实验 2** 通过与平坦调度(即文档随机均匀分布)在不同关键字条件下的比较,评估本文提出的调度算法对访问时间的优化效果。从图 4 可以看出,相比于平坦调度,本文提出的调度算法在访问时间方面占优,AU 数据集和 LA 数据

集的访问时间分别缩短了 10% 和 1% ~ 4%。由于关键字数目较少,AU 数据集文档调度产生的聚类效果相对更显著。随着关键字的数目增多,满足查询的文档数目会减少,访问时间的优化更加明显。

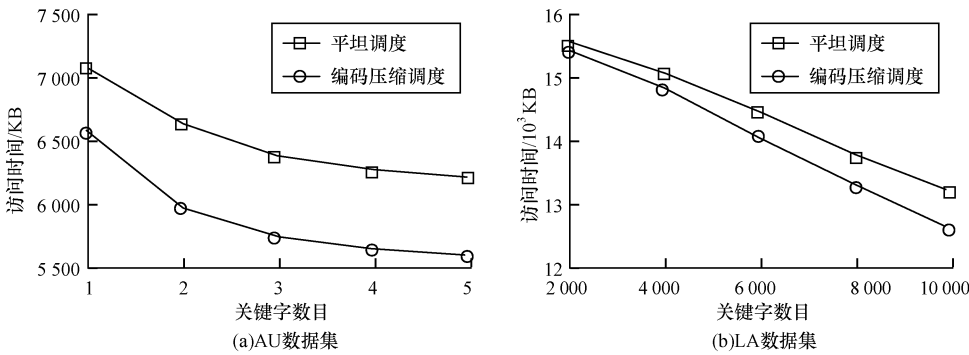


图 4 压缩前后文档调度效果比较

**实验 3** 比较不同的索引方法在不同文档数目和不同关键字个数条件下的性能。图 5 显示了随着文档数的增加,查询的访问时间和调谐时间会呈上升趋势,原因是读取的索引和文档数据的增长导致广播周期变长。而编码压缩索引在访问时间方面一直都优于原始的倒排索引和哈希索引,因为索引大小的大幅下降使得访问时间有了明显的减少。另一方面,编码压缩索引的调谐时间明显低于其他方法,

因为倒排表被压缩,读一个关键字所需下载的索引数据包数目减少,同时经过调度能降低文档数据包数量。从图 6 可以看出,随着关键字数目上升,平均访问时间和调谐时间都会下降,因为满足查询的文档数目会逐渐减少。在不同的关键字数目的情况下,编码压缩索引访问时间的表现有较大的优势。文档类型的原因,AU 数据在调谐时间方面相比大幅下降,LA 数据集优化效果稍差。

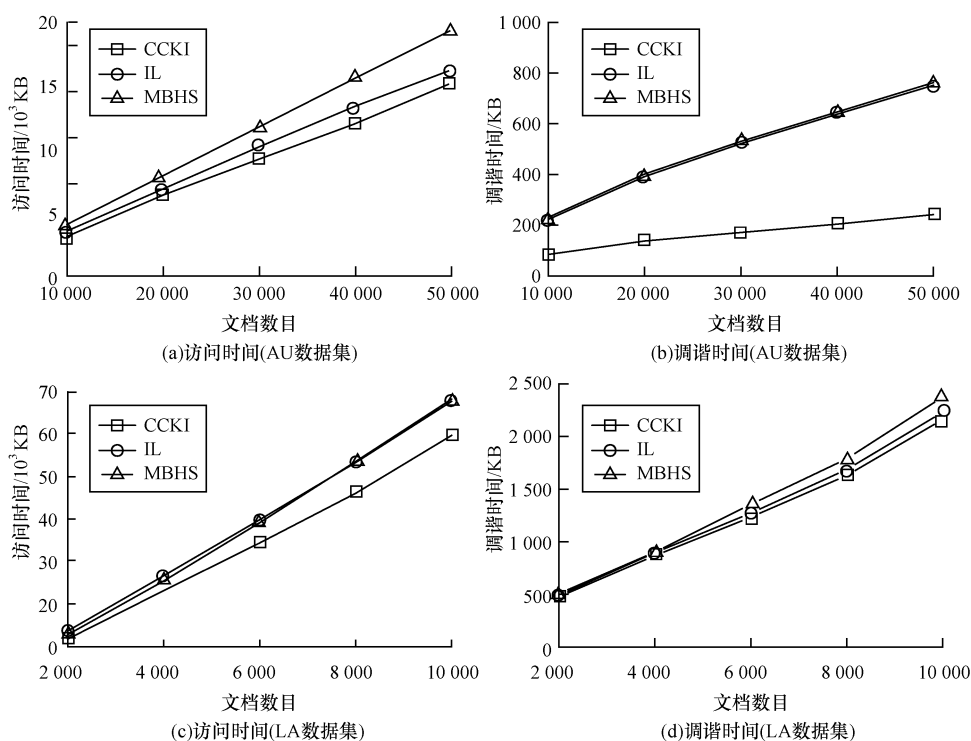


图5 不同文档数目时各调度方法的性能比较

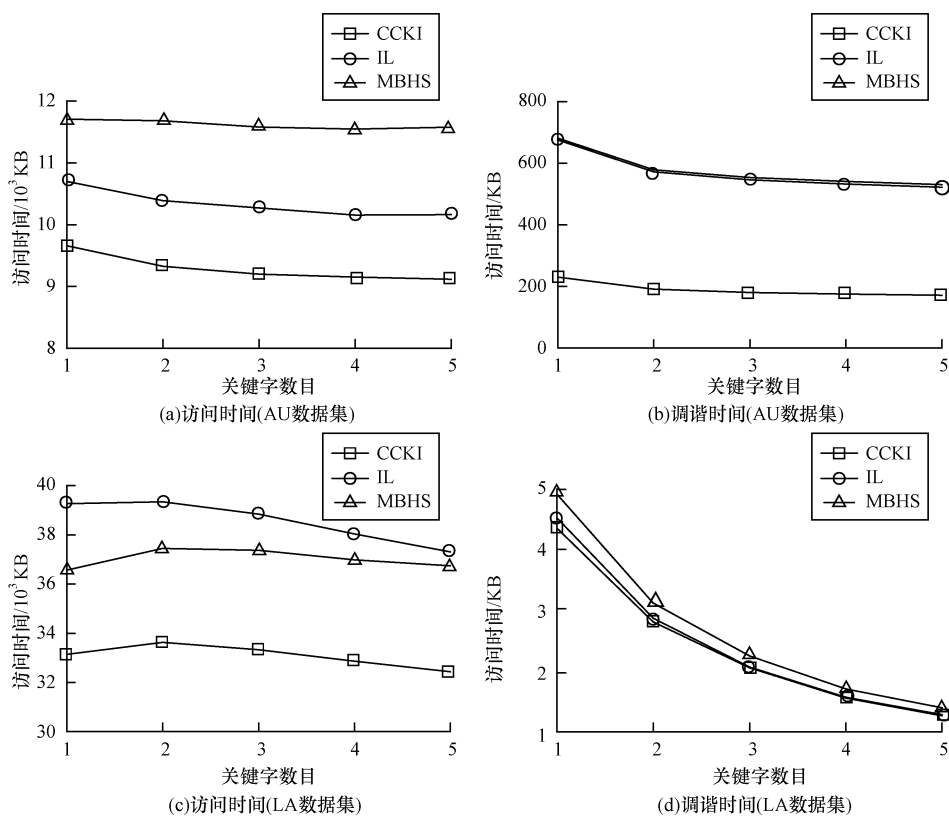


图6 不同关键字数目时各调度方法的性能比较

## 5 结束语

本文研究了无线广播环境下关键字检索的问题,提出了一种新型的关键字索引结构——编码压

缩关键字索引。关键字检索在过去已经有了大量的研究和发展,但是无线环境的特殊性导致了传统的方法不能适用于数据广播,而目前提出的倒排表索引和基于哈希的索引还无法解决索引结构过大的问

题。本文提出的关键字索引,针对倒排表进行编码压缩,缩减索引结构,并结合索引结构对文档进行调度,使访问时间和调谐时间方面的性能都得到提升。最后通过真实数据的模拟实验,验证了本文提出的关键字索引能够对以倒排表为基础的索引结构进行明显缩减,降低访问时间。下一步工作是将该索引结构从布尔模型扩展到其他信息检索模型,使其具有更广的适用性;另一方面,在周期广播中进一步改进纯文本文件的调度方法,提高系统效率。

### 参考文献

- [1] Lee X J, Hu D L, Lee Q, et al. Data Broadcast. [M]// Stojmenović I. Handbook of Wireless Networks and Mobile Computing. [S. l.]: John Wiley & Sons, 2002.
- [2] Ku W S, Zimmermann R, Wang H. Location-based Spatial Query Processing with Data Sharing in Wireless Broadcast Environments [J]. IEEE Transactions on Mobile Computing, 2008, 7(6): 778-791.
- [3] Tomasic A, Garcia-Molina H, Shoens K A. Incremental Updates of Inverted Lists for Text Document Retrieval [J]. SIGMOD Record, 1994, 23(2): 289-300.
- [4] Scholer F, Williams H E, Yiannis J, et al. Compression of Inverted Indexes for Fast Query Evaluation [C]// Proceedings of the 21st ACM SIGIR Conference on Research and Development in Information Retrieval. [S. l.]: ACM Press, 2002: 222-229.
- [5] Zobel J, Moffat A. Inverted Files for Text Search Engines [J]. ACM Computing Surveys, 2006, 38(2): 1-56.
- [6] Zhang J, Suel T. Optimized Inverted List Assignment in Distributed Search Engine Architectures [C]// Proceedings of International Parallel and Distributed Processing Symposium/International Parallel Processing Symposium. [S. l.]: IEEE Press, 2007: 1-10.
- [7] Chung Y, Yoo S, Kim M H. Energy- and Latency-efficient Processing of Full-text Searches on a Wireless Broadcast Stream [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(2): 207-218.
- [8] Yang Kai, Shi Yan, Wu Weili, et al. A Novel Hash-based Streaming Scheme for Energy Efficient Full-text Search in Wireless Data Broadcast [C]// Proceedings of the 16th International Conference on Database Systems for Advanced Applications. Hong Kong, China: [s. n.], 2011: 372-388.
- [9] Acharya S, Alonso R, Franklin M J, et al. Broadcast Disks: Data Management for Asymmetric Communications Environments [C]// Proceedings of International Conference on Management of Data. Melbourne, Australia: [s. n.], 1995: 199-210.
- [10] Chung Y, Kim M H. QEM: A Scheduling Method for Wireless Broadcast Data [C]// Proceedings of the 14th International Conference on Database Systems for Advanced Applications. Brisbane, Australia: [s. n.], 1999: 135-142.
- [11] 王丰亮,吕卫锋,诸彤宇,等. 基于贪心策略的多信道数据广播调度算法 [J]. 计算机工程, 2011, 37(12): 179-181.
- [12] Qin Y, Wang H, Sun L. Cluster-based Scheduling Algorithm for Periodic XML Data Broadcast in Wireless Environments [C]// Proceedings of IEEE International Conference on Advanced Information Networking and Applications. Ginowan City, Japan: IEEE Press, 2011: 855-860.
- [13] 吴晶晶,毛鼎鼎,朱良,等. 基于文档合并的XML无线数据广播调度算法 [J]. 计算机工程, 2011, 37(14): 31-33.
- [14] Garey M R, Johnson D S. Computers and Intractability: A Guide to the Theory of NP-Completeness [M]. San Francisco, USA: [s. n.], 1979.
- [15] Applegate D, Bixby R, Chvátal V, et al. Concorde: A Code for Solving Traveling Salesman Problems [EB/OL]. [2013-10-11]. <http://www.tsp.gatech.edu/concorde.html>.
- [16] Rocha-Junior J B, Nørsvåg K. Top-k Spatial Keyword Queries on Road Networks [C]// Proceedings of the 15th International Conference on Extending Database Technology. [S. l.]: ACM Press, 2012: 168-179.
- [17] Los Angeles Times [EB/OL]. [2013-10-11]. <http://www.latimes.com/>.

编辑 金胡考

(上接第74页)

- [8] Cadambe V R, Jafar S A. Interference Alignment and the Degrees of Freedom for the K User Interference Channel [J]. IEEE Transactions on Information Theory, 2008, 54(8): 3425-3441.
- [9] Peters S, Heath R W. Interference Alignment via Alternating Minimization [C]// Proceedings of ICASSP'09. [S. l.]: IEEE Press, 2009: 159-167.
- [10] Reingold E M, Nievergelt J, Deo N. Combinatorial Algorithms: Theory and Practice [M]. Englewood Cliffs, USA: Prentice-Hall, 1977.
- [11] Gou Tiangao, Jafar S A. Degrees of Freedom of the K User  $M \times N$  MIMO Interference Channel [J]. IEEE Transactions on Information Theory, 2010, 56(12): 6040-6057.
- [12] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by Simulated Annealing [J]. Science, 1983, 220(4598): 671-680.

编辑 索书志