

基于图的最短路径跨域数据交换实现

王 茜,李安颖,葛 新,王 浩

(西安未来国际信息股份有限公司,陕西 西安 710063)

摘 要:最短路径查找的效率决定了跨域数据交换的效率。针对通道较少($e \ll n(n-1)$)的跨域数据交换最短路径查找的问题,文中实现了一种基于图的最短路径查找方法。设计了域标识模型、域表和通道表,建立了域表与通道表的关系模型,根据面向对象的方法基于邻接表存储结构构造了域及通道的邻接表。基于深度优先搜索遍历原理,定义邻接表对象、路径集合,记录域访问历史、路径长度,以递归的方式实现了跨域最短路径的查找。实现了电子政务跨域数据交换时域间最短路径的查找,证实了文中方法的有效性。

关键词:跨域;数据交换;最短路径;邻接表;深度优先搜索遍历

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2013)12-0059-03

doi:10.3969/j.issn.1673-629X.2013.12.014

Realization of Graph-based Cross-domain Data Exchange with Shortest Path

WANG Qian, LI An-ying, GE Xin, WANG Hao

(Xi'an Future International Information Co., Ltd., Xi'an 710063, China)

Abstract: The efficiency of the shortest path searching determines the efficiency of cross-domain data exchanging. For the problem of the shortest path searching of cross-domain data exchange with fewer channels ($e < n(n-1)$), a graph-based shortest path searching method is proposed. According to the object oriented approach, after the design of domain identification model, domain table, channel table, the establishment of the relationship model between the domain table and channel table, domain and the channel adjacency list are constructed based on the adjacency list storage structure. Based on the depth-first search principle, adjacent list object and path set are defined, domain access history and the length of path are recorded, and cross-domain shortest path search method is achieved in a recursive way. The effectiveness of the proposed method is confirmed by the realization of the shortest path search of e-government cross-domain data exchange.

Key words: cross-domain; data exchange; shortest path; adjacency list; depth-first search

0 引言

各业务系统间的互联互通、信息共享、业务协同是电子政务深入发展过程中最迫切需要解决的问题^[1]。解决这些问题的关键之一在于如何在各系统间进行有效的数据交换和共享。数据交换要实现域内以及域间的信息交换,文中所指域是指相互关联的一组资源的集合,这种关联如控制、管理、授权、安全等。

随着域数量的增加,跨域最短路径的查找决定了数据交换的效率。最短路径问题是图论研究中的一个经典算法问题,旨在寻找图中两节点间的最短路径。最短路径算法有广泛的应用领域^[2-4]。常用的方法有

存在负权边的单源最短路径 Bellman-Ford 算法^[5];解决有向图中单个源点到其他顶点的最短路径问题的迪科斯彻算法(Dijkstra's algorithm)^[6];用来找出每对顶点之间的最短距离的 Floyd-warshall 算法^[7];求每对顶点间的最短路径的 Johnson 算法等^[8]。

跨域数据交换要求在不同的行政管理域或单位之间进行数据交换,在考虑安全、行政管理等因素下,网络中的域并不是任意两个域之间都有路段存在,往往是以节点数(N)较大而路段数相对完全图的路段数小很多的稀疏图($e \ll n(n-1)$)存在。鉴于此,文中基于邻接表数据结构提出一种适合于跨域数据交换最短路径查找的算法,并实现此路径的跨域数据交换。

收稿日期:2013-02-19

修回日期:2013-05-22

网络出版时间:2013-08-28

基金项目:陕西省自然科学基金(2012JM7017)

作者简介:王 茜(1966-),女,博士,高级工程师,研究方向为电子政务。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130827.1430.004.html>

1 图及其表示与搜索

假设一个图 $G=(V,E)$ 有 n 个顶点和 e 条边,顶点编号为 $1,2,\cdots,n$ 。若图 G 的边集 $E(G)$ 中为有向边,则称此图为有向图;如边集 $E(G)$ 为无向边,则称此图为无向图。边带有权的图称有权图;边不带权的图称无权图^[9-10]。

1.1 邻接表

图的表示方法有多种,邻接表是图的表示的一种常用的方法,可用于有向图也可用于无向图,可用于有权图也可用于无权图。

邻接表(adjacency list)是对图中的每个顶点建立一个邻接关系的单链表,并把它们的表头指针用向量存储的一种图的表示方法^[11]。

图 G 的邻接表由 n 个单链表组成,它们分别对应于 G 的 n 个顶点。第 i 个链表由与顶点 i 邻接的顶点(若 G 是无向图)或顶点 i 邻接到的顶点(若 G 是有向图)构成,称为边节点。边节点包含三个域:

- 一是邻接点域(adjvex),用于存储顶点 v_i 的一个邻接顶点 v_j 的序号 j ;
- 二是权域(weight),用于存储边 (v_i,v_j) 或 (v_j,v_i) 上的权;
- 三是链域(next),用于链接 v_i 邻接表中的下一个节点。

其中邻接点域和链域是必不可少的,若是无权图可以省去权域^[12]。

邻接表中,表头向量需要占用 n 个或 $2n$ 个指针存储空间,所有边节点需要占用 $2e$ (对于无向图)或 e (对于有向图)个边节点空间,其空间复杂度为 $O(n+e)$,这种存储结构适用于稀疏图。邻接表数据结构已被证明是网络表达中最有效率的数据结构,在最短路径算法中得到了广泛应用,如图 1 所示。

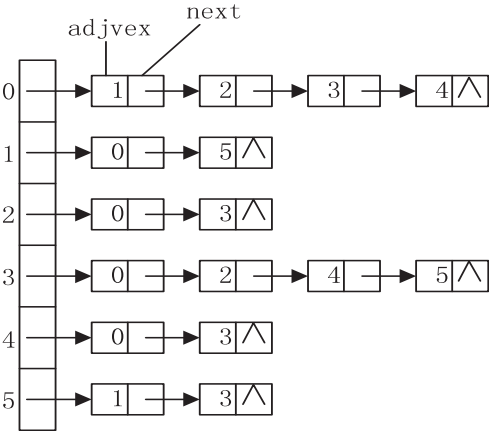


图 1 邻接表

1.2 深度优先搜索遍历

深度优先搜索(depth-first-search)遍历是图搜索算法的一种,是一个递归过程,其遍历过程为:从图中

某个节点 v_i 出发,在访问该节点 v_i 之后将其标记为已访问过,然后从 v_i 的任一个未被访问过的邻接点出发进行深度优先搜索遍历,当 v_i 的所有邻接点均被访问过时,则退回到上一个顶点 v_k ,从 v_k 的另一个未被访问过的邻接点出发进行深度优先搜索遍历,直到退回到初始点并且没有未被访问过的邻接点为止。对邻接表进行深度优先搜索遍历时,需要扫描邻接表中的每个节点,时间复杂度为 $O(e)$,空间复杂度为 $O(n)$ 。

2 跨域数据交换最短路径查找

2.1 域标识模型

域标识用于唯一地标识一个域,关键字为 mark。域标识存储在表 globaldomain 中。

表 globalchannel 用于存储通道信息,外键 leftMark,rightMark 表示数据交换的源域和目标域。

图 2 为域与通道数据库模型。

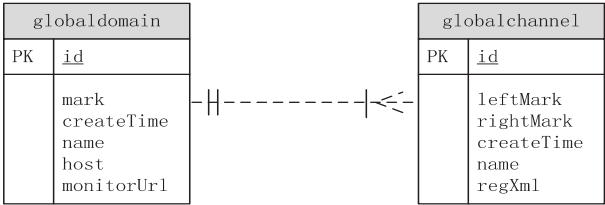


图 2 域与通道数据库模型

2.2 建立邻接表

通过面向对象语言 JAVA 实现域邻接表模型的建立,图 3 为类图。

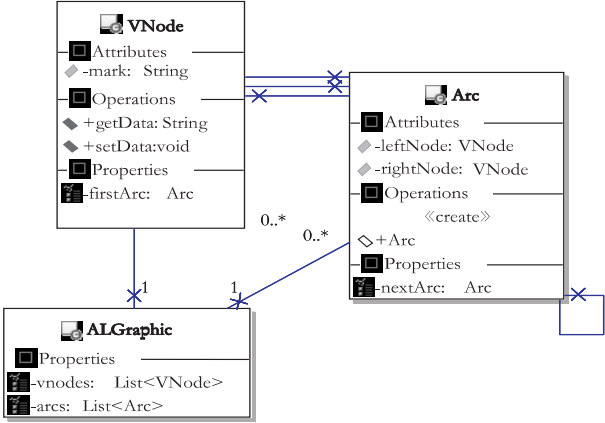


图 3 节点、边、域邻接表类图

VNode 类定义域节点,其中 data 字段表示域信息,firstArc 字段表示节点的第一个边。

Arc 类定义域节点间的边(通道)信息,leftNode 表示边的左节点,rightNode 表示边的右节点,nextArc 表示 startNode 的下一条边。

ALGraphic 类定义所有域的邻接表,vnodes 字段表示所有域节点信息,arcs 字段表示域节点的所有边信息。

ALGraphic 类中实现了邻接表的建立,建立过程如

图 4。

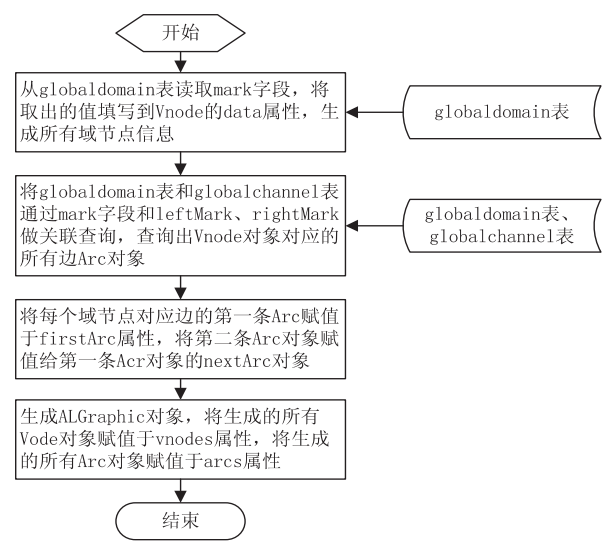


图 4 邻接表建立过程

2.3 查找最短路径

邻接表的深度优先搜索遍历的实现：

```
private void pathAll( ALGraphic g, VNode start, VNode end,
Object[] pathData, int distance)
//从初始点开始进行深度优先搜索遍历
{ this. enterGraphicNode( start); //初始化当前节点,设置其为已被访问过
distance++; //记录路径长度
pathData[ distance ] = start. getData(); //添加已经访问过的节点
//若起点等于终点,则搜索到一条结果,将结果记录到结果路径集合中
if( start. equals( end)) {
afterFindPath( pathData, end);
//定义起点的邻接边
Arc arc = start. getFirstArc();
//如果边存在
while( arc != null) {
VNode n = g. getNode( arc. getEndNo());
//如果下一个节点未被访问过,则以该节点为起点进行递归调用
if( this. visited[ n. getNo() ] == 0) {
pathAll( g, n, end, pathData, distance);
//如果下一个节点已经访问过,则开始搜索另一条路径
arc = arc. getNextArc();
this. leaveGraphicNode( start); //将终点标识为未搜索过,以便开启另一条路径的查询
}
}
}
//根据以上的设计,若要查找 0 到 3 的最短路径,根据图 1,其搜索过程为:
1) 访问 0,设置其为已被访问过,记录路径长度为 1;
```

邻接点 1、2、3、4,根据邻接表的顺序,先访问 1 节点。判断 1 未被访问过,则从 1 进行递归调用;

4) 直到访问到 3,3 为终点,得到第一条路径 0-1-5-3,路径长度为 4;

5) 将终点 3 标识为未被访问过,开启另一条路径的查询;

6) 重复步骤 1) ~ 5),查到第二条路径 0-2-3,路径长度为 3 小于 4,则当前最短路径为 0-2-3;

7) 重复步骤 1) ~ 5),查到第三条路径 0-3,路径长度为 2 小于 3,则当前最短路径为 0-3;

8) 重复步骤 1) ~ 5),查到第四条路径 0-4-5,路径长度为 3 大于 2,则当前最短路径仍为 0-3;

9) 至此,所有节点都被访问过,搜索结束,最短路径为 0-3。

3 电子政务跨域数据交换最短路径查找实现

文中描述的跨域数据交换最短路径查找方法已经在电子政务跨域数据交换中实现,用此方法可以快速实现电子政务跨域数据交换时域间最短路径的查找从而实现数据的交换共享。

图 5 为电子政务域图,各个域之间建立双向通道,通道无权重,图 6 为电子政务域的邻接表,要求查找西安域到榆林域的最短路径。根据文中给出的算法可以快速得出西安域到榆林域的最短路径为:西安域-宝鸡域-榆林域。

```
graph TD
    X[西安域] --- B[宝鸡域]
    X --- H[汉中域]
    B --- H
    B --- A[安康域]
    B --- Y[榆林域]
    H --- A
    A --- Y
    Y --- S[商洛域]
    Y --- B
    Y --- X
```

图 5 电子政务域图

域名	邻接域及方向
西安	宝鸡, 汉中
宝鸡	西安, 汉中, 安康, 榆林
汉中	西安, 宝鸡, 安康
安康	汉中, 宝鸡, 榆林
延安	宝鸡
商洛	榆林
榆林	宝鸡, 汉中, 安康, 商洛, 延安

图 6 电子政务域的邻接表

一次访问 Memcached Manager 所需要的时间要比直接读数据库长,因为第一次读取需首先读取数据库,然后将数据存入缓存系统,第二次访问则直接从 Memcached 中获取。

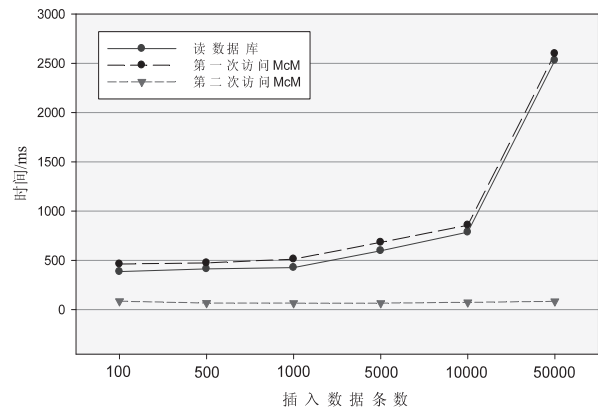


图 2 不同的查询方式所需的时间

经过上边的结果分析,在实际的应用系统中,相对于传统的缓存系统,文中提出的方案同样可以有效提升系统的运行效率,尤其是在大数据量的情况下,其优势更加明显。

5 结束语

文中首先分别从 Memcached 应用特征和 Web 应用体系结构特征两方面分析了现有 Memcached 系统的缺点,从而提出了一种缓存资源集中管理分配方法和多应用共享方案,并介绍相应的 Memcached Manager 应用系统。实验结果表明,相对于传统 Web 开发方式,文中提出的方案能够获得很高的性能加速,同时可以有效提高开发人员的开发效率。

4 结束语

文中基于邻接表存储结构和深度优先搜索遍历,以面向对象的方法构造了域邻接表和搜索算法,给出了详细的运算流程,并且在电子政务跨域数据交换实际场景中实现了该方法,为以稀疏图形式存储的电子政务域的跨域数据交换最短路径查找提供了一条新途径。

参考文献:

[1] 邓磊,吴健,张昌利,等.电子政务中跨域可信数据交换模型设计与实现[J].计算机工程,2007,33(12):4-6.

[2] Cherkassky B V,Goldberg A V,Radzik T. Shortest paths algorithms:theory and experimental evaluation[J]. Mathematical programming,1996,73:129-174.

[3] Lu Feng,Zhou Chenghu,Wan Qin. An optimum path algorithm for traffic network based on hierarchical spatial reasoning[J].

参考文献:

[1] Memcached: High-performance, distributed memory object caching system[EB/OL]. 2011. <http://memcached.org/>

[2] Vaidyanathan K,Narravula S,Balaji P,et al. Designing efficient systems services and primitives for next-generation data centers[C]//Proc of workshop on NSF next generation software (NGS) program. [s. l.]:[s. n.],2007.

[3] Tanenbaum A S,Steen M V. Distributed systems:Principles and paradigms[M]. 2nd ed. [s. l.]:Pearson Prentice-Hall,2006.

[4] Petrovic J. Using Memcached for data distribution in industrial environment[C]//Proceedings of the third international conference on systems (ICONS). [s. l.]:[s. n.],2008:368-372.

[5] Menace D A. Scaling Web sites through caching[J]. IEEE Internet computing,2003,7(4):86-89.

[6] 秦秀磊,张文博,魏峻,等.云计算环境下分布式缓存技术的现状与挑战[J].软件学报,2013,24(1):50-66.

[7] 陈康,郑纬民.云计算:系统实例与研究现状[J].软件学报,2009,20(5):1337-1348.

[8] Karger D,Lehman E,Leighton T,et al. Consistent hashing and random trees: Distributed cachine protocols for relieving hot spots on the World Wide Web[C]//Proc of 29th annu ACM symp on theory of computing. [s. l.]:[s. n.],1997:654-663.

[9] 林海略,韩燕波.多租户应用的性能管理关键问题研究[J].计算机学报,2010,33(10):1881-1895.

[10] 王伟,曾国荪.基于 Bayes 认知信任模型的 MANETs 自聚集算法[J].中国科学 E 辑,2010,40(2):228-239.

[11] 王伟,曾国荪.一种基于 Bayes 信任模型的可信动态级调度算法[J].中国科学 E 辑,2007,37(2):285-296.

Geo-spatial information science,2000,3(4):36-42.

[4] 陆锋.最短路径算法:分类体系与研究进展[J].测绘学报,2001,30(3):269-275.

[5] 宫恩超,李鲁群.基于 Bellman-Ford 算法的动态最优路径算法设计[J].测绘通报,2011(8):26-28.

[6] 鲍培明.距离寻优中 Dijkstra 算法的优化[J].计算机研究与发展,2001,38(3):307-311.

[7] Höfner P,Möller B. Dijkstra,Floyd and Warshall meet Kleene[J]. Formal aspects of computing,2012,24(4-6):459-476.

[8] 王银燕,余镇危,曹怀虎,等.基于二度量的单播最短路径算法[J].计算机工程,2007,33(5):89-90.

[9] 许国志.系统科学[M].上海:上海科技教育出版社,2000.

[10] 徐振宁,黄凯歌,张维明,等. Ontology 建模方法研究[J]. 计算机科学,2002,29(1):68-71.

[11] 徐孝凯.数据结构[M].北京:电子工业出版社,2004.

[12] 温晓磊.基于混合算法的最短路径优化算法[J].天津理工大学学报,2009,25(1):37-40.

基于图的最短路径跨域数据交换实现

作者：[王茜](#)，[李安颖](#)，[葛新](#)，[王浩](#)，[WANG Qian](#)，[LI An-ying](#)，[GE Xin](#)，[WANG Hao](#)

作者单位：[西安未来国际信息股份有限公司, 陕西 西安, 710063](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

ISTIC

年，卷(期)：2013(12)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjfz201312014.aspx