



一个综合 IP 地址查询系统的设计与实现

蔡少敏, 丁伟, 张维维

(东南大学计算机科学与工程学院, 南京, 210000)

摘要: 为向中国教育和科研网华东地区网络中心的用户提供准确、详尽且更符合需要的IP地址归属信息查询, 在核实其身份的基础上, 以异步的方式提供相应的服务, 同时按常规方式, 为普通用户提供缩减的IP地址归属查询服务。本系统是一个采用B/S设计模式, 基于LAMP的信息查询管理系统。在本系统中, 主要实现由whois数据源获取IP地址归属表, 并基于该IP地址归属表实现用户登录模块、用户管理模块、用户查询模块、系统管理模块, 并通过socket通信来实现IP地址发布系统对外公开提供服务等。本文详细介绍了该综合IP地址查询系统的功能需求、系统设计、实现细节和开发过程及方法。

关键词: IP 地址; 地理位置; ISP; 具体单位; CERNET 用户; MySQL 数据库

Design And Implementation For An Integrated IP Address Searching System

Cai Shaomin, Ding Wei, Zhang Weiwei

(School of Mechanical Engineering, Southeast University, Nanjing, 210000)

Abstract: The thesis is to provide more accurate - detailed IP address information for all the China Education and Research Network center in northeastern regional networks. It can meet the needs of IP address information query services. What's more, it also provides some queries in asynchronous delivery for some management related to Education network, but they need to be identified. For most user, the thesis provides basic IP address location research. We use the B/S patterns and LAMP, to achieve the login module of different users, to realize the searching module, and the system management module. The paper implements the IP address information table from the source whois, and analyzes the functional needs of the system - the system design and implementation. Also, it briefly introduces the development process and methods we used.

Key words: IP address; Location; ISP; Specific unit; CERNET users; MySQL Database

CERNET华东地区网络中心在网络运行与网络管理、网络安全有关的科研工作中, 存在大量的与IP归属有关的需求。目前, 互联网上有许多在线提供IP地址归属信息服务的商业网站, 比如www.ip.cn、www.ip138.com、ipseeker.cn等, 也有Whois这样通过命令接口实现对域名信息的查询。虽然它们都能够提供IP地址归属信息的查询, 但是由于这些网站均以匿名方式提供服务, 所以在功能方面会受到很大限制, 比如www.ip.cn支持的查询条件比较单一, 只能根据IP地址查其归属, 并且不能支持大量的IP地址的同时查询, 也不支持以IP地址

博士研究生, E-mail: wwzhang@njnet.edu.cn

以外的信息作为关键字进行查询。其他网站也会有限制每天的查询次数、查询到的归属信息不够详细全面等问题。

为了给网络中心科研和运行工作提供全方位的IP地址归属服务, 我们基于whois数据库构建了一个综合IP地址归属查询服务系统, 该系统以任意表项缺省和组合方式构造查询条件, 并以灵活的方式返回表所有符合条件的IP地址。

本系统在为网络中心有关的科研工作提供全方位的地址归属服务的基础上, 还可以逐步向江苏教育网中有类似需要的其他实名用户开放。

作者简介: 蔡少敏, (1987-), 女, 硕士研究生, E-mail: smcai@njnet.edu.cn; 丁伟, (1962-), 女, 教授, E-mail: wding@njnet.edu.cn; 张维维, (1983-), 男,

工作方案和系统的功能

为实现上述目标, 本论文的工作分成 2 个方



面展开：一是全局 IP 地址表的获取，二是基于该地址表完成一个实际的查询系统。更具体而言，第一项工作的目标是建立一个全局 IP 归属表，第二项工作是基于这张表设计并实现一个综合查询服务系统。

全局 IP 地址归属表的获取我们采用基于 whois 原始数据库查询注册记录，通过字符串匹配算法进一步提取其归属信息的方式完成，该表的格式如表 1 所示。

表 1 IP 地址·属表

IP	continent	country	region	city	isp	unit
----	-----------	---------	--------	------	-----	------

查询系统基于上面的全局 IP 归属表，采用 B/S 结构实现。在功能方面，该系统对普通用户支持 IP 地址归属信息；对 CERNET 用户，该系统提供了更加详细的查询方式，可以将该表的任何表项、或者联合几个表项作为查询条件（各项查询条件之间的关系可以是与关系，也可以是或关系），并返回所有符合条件的记录，同时支持用户以 excel 格式进行下载返回结果。详见系统功能图（图 1）。

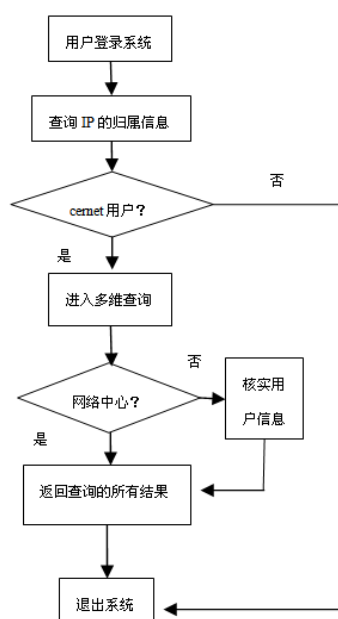


图 1 系统功能图

全局 IP 地址归属表的建立

本系统的首要任务是通过获取的 IP 地址归属的数据源，来构建一个比较完整且准确的 IP 地址归属表。该

表的归属信息包括地理归属（所属国家、省、市）和逻辑归属（所属 ISP、POP 主节点、单位名称）。数据源的获取可以采用基于数据挖掘的方法和基于时延测量的方法。基于数据挖掘的方法包括基于 DNS LOC^[1]、whois 数据库^[1]、GeoTrack^[2]、GeoCluster^[2]、和基于 web 数据挖掘和启发式推断^[3]。基于时延测量的方法包括 GeoPing^[4]、基于约束的地理定位^[5]、基于拓扑的地理定位^[6]。

1.1 获取数据源

本系统首先采用基于数据挖掘中的查询 whois 数据库注册记录的方法。whois 数据库是一种由 DDN 信息中心维护的关于用户、主机系统、网络和域的 Internet 数据库。从 whois 数据库可以获得每个 IP 地址的相关地理位置等许多信息。如 whois 202.112.23.35，可获取信息如下：

inetnum: 202.112.0.0-202.112.31.255

netname: CEPBKB-CN

descr: China Education and Research Network
Center

descr: Tsinghua University

descr: Beijing, 100084

country: CN

admin-c: CER-AP

tech-c: CER-AP

remarks: origin AS4538

remarks: service provider

changed: hn-changed@net.edu.cn 19960211

mnt-by: MAINT-CEPNET-AP

status: ASSIGNED NON-PORTABLE

source: APNIC

将查询到的每个 IP 地址/段的信息选择系统所需要的逐条写入文件 IP_location_file_webwhois 中。

文件包含信息：

IP 地址/前缀 | 洲 | 国家 | 具体单位 | info

其中洲、国家和具体单位是可以直接提取到的而 info 是间接提取到的关于 IP 归属的模糊信息。

一个 IP 地址/前缀可能需要多次 whois 数据库查询，其基本算法如图 2 所示：

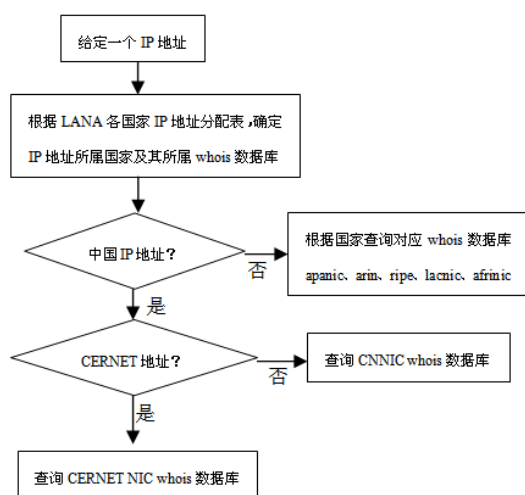


图 2 获取数据源算法

1.2 构建 IP 归属表

然后, 由上面所获取的数据源文件 IP_location_file_webwhois, 通过字符串匹配算法进一步提取其归属信息, 从而构建 IP 地址归属表。在做词法分析时, 需要用到一些基本信息, 分别是用来记录国家、该国家下的省市及 ISP、具体单位等。读取这些信息构建数据结构, 并用数组存放, 以便在字符串匹配时使用。

主要的数据结构 Record, 来记录每一个 IP 地址/前缀对应的所有信息:

```
struct Record
{
    char IPAddr[19];    //IP 地址加前缀
    char RIR[8];        //所属 RIR
    char country_s[3];  //国家名缩写
    char region[REGION_LEN]; //省名
    char city[CITY_LEN]; //城市名
    char ISP[ISP_LEN];  //ISP 名
    char unit [UNIT_LEN]; //具体单位或学校名
} record;
基本算法:
```

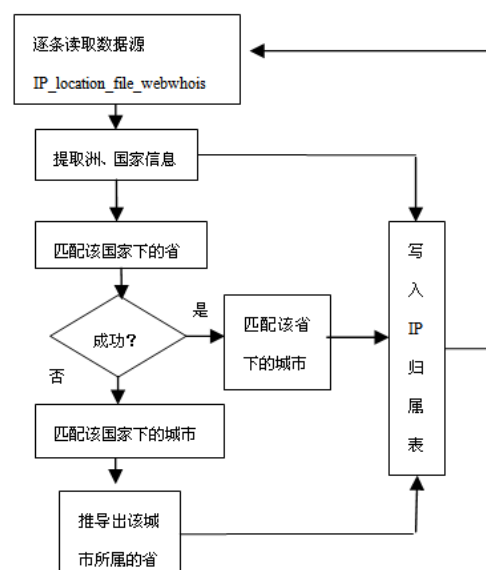


图 3 构建 IP 归属表算法

1.3 IP 归属表的维护

由于本系统的一切运转都是基于该 IP 归属表的, 因此对于该 IP 归属表的维护尤为重要。

首先, 为了防止 IP 归属表受到攻击或窃取, 将其放置于内网服务器中, 只提供一个端口来窃听外网上提供服务的一个服务器的请求。这其中使用 socket 通信来实现^[7], 将在 4.4 中具体介绍。

其次, 由于用以上方法获得的 IP 归属表会存在地理定位误差, 因此需要进一步提高该表的准确性。本文采用测量 RTT 来找出那些地理定位错误的 IP 地址。通过测量分别在两个域内的 IP 地址之间报文发送接收的 RTT, 比较找出与其他 RTT 差别非常大的 IP 地址, 从而来进行勘误。



查询系统的结构设计

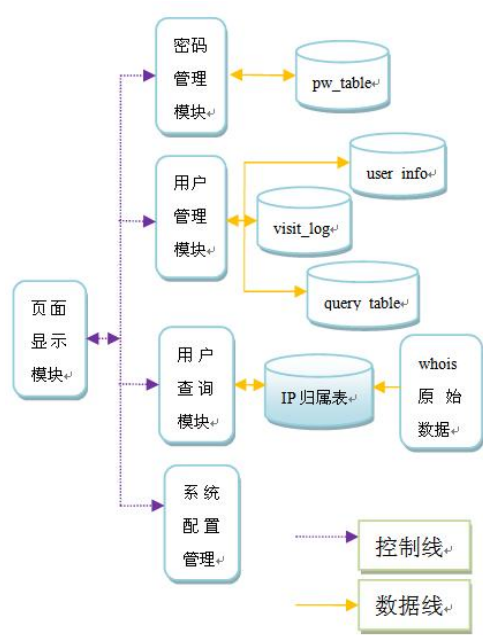


图4 系统结构图

页面显示模块：负责接收用户输入请求，返回查询结果

密码管理模块：负责每天更新一个八位随机数密码，密码的验证及邮件的发送

用户管理模块：负责对用户进行管理，判断用户模式。记录用户信息、访问时间及其具体查询条件信息等，CERNET用户的基本信息，并将该信息发送至管理员邮箱

用户查询模块：根据用户请求，查询IP归属表，将查询结果提供给页面显示模块

系统配置管理：对系统进行配置和维护管理

1) 更新IP归属表

2) 系统出错时的恢复

3) 用socket实现内外网服务器的通信

查询系统的实现

1.4 密码管理模块

每天凌晨，系统将进行一次密码更新，并将该随机密码存入数据库，同时发送至管理员的邮箱，以防系统崩溃时查询使用。

密码的更新实现技巧：每天，当第一个用户访问该系统时，系统将该用户此时的登录日期与数据库中password_log中最后一个密码的生成日期作比较，由于是当日第一次登录，此时必然不同，则立即生成当日新的密码，同时将该密码发送至管理员邮箱

对于每个用户，当天若是首次登录该系统，则需通过邮箱获取当日密码，此时将该用户的ip地址、邮箱以及登录时间记录到数据库。

实现将当日密码发送至客户端邮箱，需要用到两个类：class Mail 和 class Mail_mine

1.5 用户管理模块

利用每个用户的IP地址不同，来区分标志该用户，并通过其来判断该用户是否为CERNET用户。

对于每一位客户，系统将该用户的基本信息及查询信息记入数据库。在非网络中心的CERNET用户使用多维查询功能时，需要记录该用户的完整的信息，以便核实后反馈查询结果。

1.6 用户查询模块

该模块是系统最主要的模块。不同的用户设有不同的查询权限。所有的用户都可以由IP地址查其归属，而只有CERNET用户可以进入系统的多维查询。对于普通用户，系统支持一次可查询最多三个IP地址的归属信息；对于CERNET用户的多维查询（如图5），可以将IP归属表的任何一项或者多项的组合作为查询条件，如查询条件：地理位置为亚洲中国江苏省南京市并且ISP为教育网，系统返回符合条件的查询结果。



请填写您的查询条件:

IP地址:

1. 2.

3. 4.

注: 支持模糊查询, 如ip地址: 202.112.23.*, 202.112.*.*

地理位置:

请选择 请选择 请选择 请选择

ISP(互联网服务提供商):

具体单位或学校:

☒ 与 ☐ 或

*该系统可根据四项条件查询, 当查询多项时, 各项默认“与”关系, 也可手动选择

其中IP地址可以查询四段, 之间默认为“或”关系。

图 5 多维查询

如上图的多维查询界面所示, 对于 IP 地址条件的输入, 系统提供了四个 IP 地址的输入, 用户可以根据需要输入一个或者多个 IP 地址作为查询条件, 本系统支持 IP 地址的模糊查询, 输入的 IP 地址可以为 202.112.23.*。对于地理位置条件的选择, 本系统提供了四级级联菜单, 用户可以大范围的查询, 即只选择某个洲; 也可以小范围的查询, 即具体到某一个城市。对于国外地理位置的选择只能具体到国家, 而中国的地理位置可以具体到一个城市。对于 ISP (互联网提供商) 的选择, 系统列出了六个选择项, 即教育网、电信、网通、移动、联通和铁通。对于具体的单位或者学校, 系统可以允许用户自己输入, 同时为了方便用户, 系统支持 google 搜索的 auto suggest 功能, 如用户欲查询上海财经大学的相关信息, 在输入框输入上海时系统会自动搜索上海的单位 and 学校, 弹出的下拉菜单中就包括上海财经大学、上海第二工业大学、上海师范大学等, 用户选择可以直接选择所需单位或学校。

以上的各个查询项, 他们之间的关系默认为“与”关系, 当然, 用户也可以手动选择“或”关系。

冲突策略: 在多维查询中, 由于将 IP 归属表中的其他表项作为查询条件, 其返回结果和简单的 IP 地址归属查询结果有所不同。因为一个 IP 地址最多对应一条查询结果, 而一个地理位置或者一个 ISP, 或者一个具体单位学校, 有可能对应很多的查询结果。所以当查询条件中含有 IP 地址

时, 为了避免 IP 地址的重复, 查询结果只选取掩码最大的那个段; 当查询条件中不含有 IP 地址时, 查询结果数量会可能比较多, 采取分页显示返回结果。

1.7 使用 socket 通信来实现内网对外公开提供服务

IP 地址发布系统的实现对象分为三种:

- 1、网络中心的工作人员, 主要进行研究工作、网络管理以及网络开发等;
- 2、CERNET 用户, 如重点实验室、图书馆需要定向提供 IP 地址综合信息服务;
- 3、普通的非 CERNET 用户

基于对以上三种对象提供服务, 本系统在对外提供该服务时, 实现的通信方案如图 6 所示:

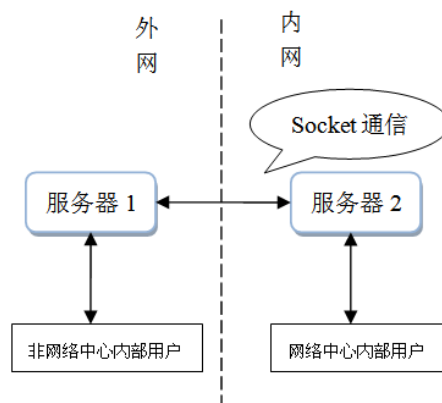


图 6 socket 通信

服务器 1——在该服务器上存放非内网用户所访问的网页, 在有用户访问时, 该服务器变成客户端, 向服务器 2 的特定端口发起连接通信, 发送服务请求报文并等待接收应答

服务器 2——在该服务器上存放供查询使用的数据库以及供网络中心内部用户访问的网页, 并时刻运行着来自服务器 1 的监听进程, 等待客户请求到达该端口。对于来自网络中心内部用户的查询请求, 直接连接服务器返回查询结果; 而对于非网络中心内部用户则需通过服务器 1 向服务器 2 发出连接请求, 发送服务请求报文, 服务器 2 通过本地的数据库查询, 向服务器 1 返回查询结果。



系统的测试

1.8 页面跳转流程图

在进入系统进行查询时，页面跳转主要分为两个流程：对于网络中心的研究人员，系统支持其多维查询，无需核实用户信息并可立即返回符合查询条件的结果；而对于普通用户，若进行多维查询，则需核实其身份，并且以异步方式返回符合查询条件的结果。

5.1.1 网络中心研究人员

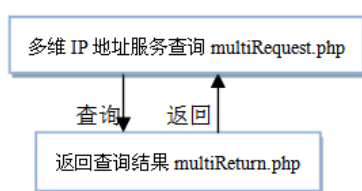


图 7 网络中心研究人员的查询流程

5.1.2 普通用户

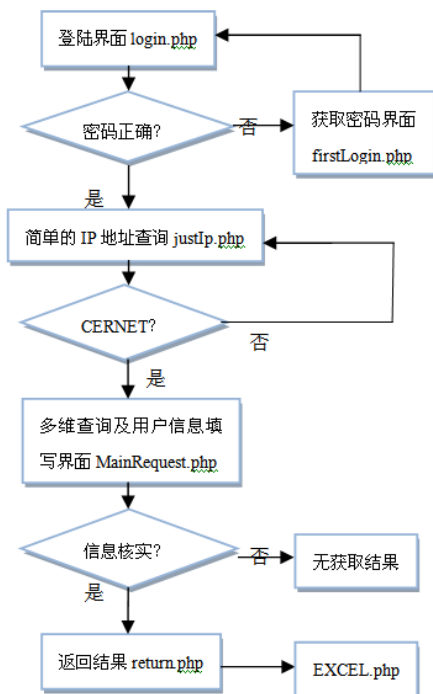


图 8 普通用户的查询流程

1.9 测试流程

- 1) 登录，输入密码，密码由数字 0~9 和字母 a~z, A~Z 随机形成，共八位。
- 2) 若是今日首次登录或者忘记当日密码，则可以点击登陆页面的“获取密码？”，进入密码获取界面。登录所输入的邮箱，获取当日密码，并返回登录页面重新登录。
- 3) 登录成功后进入简单的 IP 地址归属查询界面，如 9 所示：



图 9 简单的 IP 归属查询界面

如查询两个 IP 地址 202.112.23.35 和 58.192.112.23，页面直接返回查询结果，如表 2 所示：

表 2 查询结果

IP 地址/前缀	洲	国家	地区	城市	ISP	子网名称
202.112.23.35	亚洲	中国	北京	北京	教育网	CERNET
58.192.112.23	亚洲	中国	江苏	南京	教育网	东南大学

4) 若用户为 CERNET 用户，且欲进入多维查询界面查询更多信息。点击如图 9 所示的“CERNET 用户？”进入多维查询界面，如图 5 所示。

输入查询条件后，要得到查询结果，需先核实用户信息及查询用途。在填写用户信息时，除了手机号码外必须全部真实填写。管理员在核实了该用户的身份和用途后，将符合查询条件的结果以 excel 格式发送到该用户的邮箱。如果该用户在填写用户信息时提供了手机号码，管理员将会同时发送一条提示短信给该用户。

5) 若是网络中心研究人员，则直接返回符合查



询条件的结果，并支持将该结果以 excel 表格的形式进行下载。

如查询条件：地理位置为亚洲中国江苏南京，ISP 为教育科研网，具体单位为东南大学。查询结果直接以表格的形式返回给用户，如图 10 所示：

您的查询结果如下:

IP地址	洲	国家	地区	城市	ISP	子网名称
58.200.5.0/24	亚洲	中国	江苏	南京	教育网	东南大学
58.200.73.0/24	亚洲	中国	江苏	南京	教育网	东南大学
58.200.74.0/24	亚洲	中国	江苏	南京	教育网	东南大学
211.65.232.0/22	亚洲	中国	江苏	南京	教育网	东南大学
202.119.144.0/22	亚洲	中国	江苏	南京	教育网	东南大学
121.249.16.0/20	亚洲	中国	江苏	南京	教育网	东南大学
121.249.0.0/20	亚洲	中国	江苏	南京	教育网	东南大学
121.248.48.0/20	亚洲	中国	江苏	南京	教育网	东南大学
58.192.112.0/20	亚洲	中国	江苏	南京	教育网	东南大学
202.119.0.0/19	亚洲	中国	江苏	南京	教育网	东南大学
121.248.0.0/19	亚洲	中国	江苏	南京	教育网	东南大学
211.65.32.0/19	亚洲	中国	江苏	南京	教育网	东南大学

图 10 查询结果

研究总结

科技一直在飞速发展，用户对IP地址的相关信息的需求也越来越高。因此需要这样的查询系统的准确度更高，覆盖范围更广，以及返回结果更为全面。而且现在IPv4的下一代互联网协议版本IPv6，正在不断发展和完善，不久的将来将取代目前被广泛使用的IPv4。

尽管本系统实现了预期功能，但是采用web whois数据库获取IP归属表，会存在以下缺点：whois数据库会存在信息陈旧且难免出现填写错误的问题；大公司常常注册同样的信息，导致地理定位出现误差；在通过字符串匹配来寻找地理归属信息时会出现错误匹配问题。这些问题，我们将在下一阶段逐一解决，尽量提高其准确率。

参考文献：

- [1] D. Moore, R. Periakaruppan, J. Donohoe, et al. Where in the World is netgeo.caida.org? [J]. In Proceedings of INET'00, 2000. 1-6.
- [2] V. N. Padmanabhan, L. Subramanian. An Investigation of Geographic Mapping Techniques for Internet Hosts. In Proceedings of SIGCOMM'01, 2001. 8-10.

- [3] Chuanxiong Guo, Yunxin Liu, Wenchao Shen, et al. Mining the Web and the Internet for Accurate IP Address Geolocations [J]. Infocom, 2009. 4-7.
- [4] V. N. Padmanabhan, L. Subramanian. Determining the geographic location of Internet hosts. In SIGMETRICS/Performance, 2001.
- [5] B. Gueye, A. Ziviani, M. Crovella, S. Fdida. Constraint-Based Geolocation of Internet Hosts. IEEE/ACM trans. Networking, 14(6), Dec 2006.
- [6] E. Katz-Bassett, J. P. John, A. Krishnamurthy, et al. Towards IP Geolocation using Delay and Topology Measurements. In Proceedings of IMC'06, 2006.
- [7] 吴祖民, 赵勇, 杨仁庆. Linux 下网络 Socket 编程技术 [J]. 通信与广播电视, 2005(2).
- [8] 聂庆鹏, 毛书鹏, 王志乐. PHP+MySQL 动态网站开发与全程实例[M]. 北京: 清华大学出版社, 2007.
- [9] 张建立. Dreamweaver 动态网站建设全程实录[M]. 科学出版社, 2009. 320-514.
- [10] 武洪萍, 马桂婷. 数据库原理及应用(SQL Server 版) [M]. 北京: 北京大学出版社, 2008.
- [11] 刘智勇. Javascript 开发技术大全[M]. 北京: 清华大学出版社, 2009.
- [12] 方晨. Dreamweaver/PHP/MySQL 动态网站建设实务 [M]. 上海: 上海科学普及出版社, 2009. 1-262.
- [13] 徐国平. 网页设计与制作教程[M]. 第二版. 高等教育出版社, 2002.
- [14] 聂小燕, 鲁才, 许文波. 美工神话(CSS 网站布局与美化)[M]. 人民邮电出版社, 2007.