

一种基于地理信息元数据标准的空间索引体系

赵鹏举¹, 刁明光², 薛涛², 李建存³

1. 中国地质大学(北京)信息工程学院, 北京 100083
2. 中国地质大学(北京)软件学院, 北京 100083
3. 中国国土资源航空物探遥感中心, 北京 100083

摘要 在分布式空间数据库网络环境下, 根据标准《地理信息元数据》, 建立属于具体地质调查信息平台的空间数据库级和空间索引图层级两级元数据结构, 并依此集成空间数据, 构建统一的空间索引数据库; 在空间索引范围内通过生成 Hilbert 空间填充曲线扫描矩阵, 得到 Hilbert 空间排列编码与空间索引信息的一维映射关系, 设计出合理、高效的空间索引算法。通过整体规划和设计, 构建了统一的运行方式、数据集成方法和系统体系结构。试验表明, 系统通过直接对关系型数据库级表之间的操作和维护, 避免了调用系统依托的 ArcGIS 开发平台下 ArcObjects 封装的相关函数, 在实现空间索引功能的同时, 提高了空间索引效率, 为地质调查信息平台建设提供了关键的空间索引体系和良好的数据共享支持。

关键词 空间索引; 空间元数据; Hilbert 空间排列; 地质调查信息; 数据共享

中图分类号 P208

文献标识码 A

文章编号 1000-7857(2010)08-0084-05

A Geospatial Index Architecture Based on Geological Information Metadata Standard

ZHAO Pengju¹, DIAO Mingguang², XUE Tao², LI Jiancun³

1. School of Information Technology, China University of Geosciences, Beijing 100083, China
2. School of Software, China University of Geosciences, Beijing 100083, China
3. China Aero Geophysical Survey & Remote Sensing Center for Land and Resources, Beijing 100083, China

Abstract According to the deployment of the spatial database in a distributed network environment, based on current geological information metadata standards, a specific geological survey information platform is established for metadata structures, including both the spatial database level and the layer of spatial index. So the spatial data are integrated, a unified spatial index database is built in the space within the framework of the index generating Hilbert space-filling curve scan matrix, to get one-dimensional mapping between Hilbert space arrangement of coding and spatial index information. A reasonable and efficient spatial index algorithm is designed for the overall planning and design. Unified operating methods are proposed for data integration, with a system architecture to improve the level of work for Geological Survey of information technology and complete data sharing and information service system. Experiments show that the system can directly access a relational database table for operation and maintenance, to avoid as much as possible the calls of the ArcGIS platform package ArcObjects correlation functions. In the realization of spatial indexing at the same time, the efficiency of spatial index is greatly enhanced. It provides a key spatial index system and a good support for data sharing for the geological survey information platform.

Keywords geospatial index; spatial metadata; Hilbert spatial ordering; geological survey information; data sharing

收稿日期: 2010-01-19

基金项目: 中国地质调查局“资源与环境遥感信息化基础平台建设”项目(1212010650602)

作者简介: 赵鹏举, 硕士研究生, 研究方向为应用数学, 电子信箱: qunpeng@163.com; 刁明光(通信作者), 讲师, 研究方向为地球探测与信息技术、软件工程, 电子信箱: dm@cugb.edu.cn

0 引言

随着网络技术和 Web GIS 的发展及应用, GIS 相关部门产生的海量空间数据均存储在分布式的各个数据库服务器中, 相关运行系统之间空间数据的合理共享成为 GIS 发展的瓶颈。如何建立一个地理空间信息统一管理和合理共享并提供数据服务的信息化基础平台, 构建网络环境下分布式空间数据的高效索引机制, 引起了相关研究者和数据服务决策者的关注。韩新春^[1]提出了一种实现元数据和数据集一体化管理的有效途径, 在对元数据和数据集进行分布式管理的同时, 协调了元数据服务和数据集服务的一致性和同步性。Tao Xue 等^[2]设计了基于标准《地理信息元数据》^[3]的元数据管理平台, 有效解决了平台应用部门之间空间数据的集成和共享问题。陈占龙等^[4]研究了分布式空间数据索引机制, 基于 R-树和散列 hash 表, 提出了一种分布式环境下面向海量空间数据的分布式索引树结构——DR-H 树, 有效提高了分布式网络环境下海量空间数据的检索性能。

文献检索及地质调查信息化工作调研表明, 海量空间数据的集成共享和分布式空间索引面临以下问题: ① 各个相关部门生产出大量可用的空间数据, 这些海量的空间数据分布存储在网络环境下的各个服务器中; ② 对所产生的空间数据的描述, 即所采用的元数据标准和格式不尽相同, 使得地理空间信息的终端用户不能快速有效地索引出所需要的目标空间知识; ③ 即使采用了统一的标准进行空间数据集成, 当多用户并发访问空间索引系统时, 大数据量的地理空间信息也会增加系统服务器的运行负担, 降低空间索引效率。针对地质调查信息平台的空间数据集成共享和分布式空间索引, 本研究设计了基于标准《地理信息元数据》的空间索引体系, 并将其应用在系统平台建设过程中。

1 地理信息元数据标准

元数据旨在为各种形态的数字化信息单元和资源集合提供规范普遍的描述方法, 帮助数据生产单位有效管理和维护数据, 利用元数据信息通过网络对数据进行查询检索和数据的传输交换。元数据为分布的、由多种数字化资源有机构成的信息体系提供整合的工具与纽带, 因此元数据是数据集成的核心技术, 也是系统集成的有利工具^[1]。

在地质调查平台的建设过程中, 数据库集成与管理主要依据的标准是《地理信息元数据》, 根据该标准进行空间数据库集成与管理, 建立元数据库, 其主要由 8 个元数据子集 (UML 包) 构成, 包括元数据信息、空间参照信息、内容信息、数据质量信息、引用信息、标识信息、负责单位联系信息、分发信息。

2 地质调查信息平台元数据结构

建设地质调查信息平台的目标是提供数据服务, 实现数据共享, 使数据在相关领域得到充分利用。数据共享的关键

是采用统一的数据库建立标准, 即建立统一的元数据的格式和元数据与空间数据之间的映射关系。

根据对地质调查信息化平台基础数据的整理、对平台服务对象数据使用情况的调研, 为使数据集成和共享工作在地质调查信息平台服务系统内部能够方便快捷地进行, 在空间索引过程中不产生无关的搜索, 提高索引效率, 本研究在《地理信息元数据》的基础上, 删除了数据质量信息、引用信息、分发信息 3 个数据子集, 添加了空间数据库信息、空间索引信息、项目信息子集。并直接在负责单位联系的信息与项目信息之间建立关联, 实现了某一工作项目与对应元数据的负责单位联系信息的统一管理, 这样负责单位联系信息只由项目信息子集调用而不需要在元数据信息中管理, 实现了数据集成信息化的统一, 减小了空间索引系统的开销。地质调查平台元数据的子集如图 1 所示。

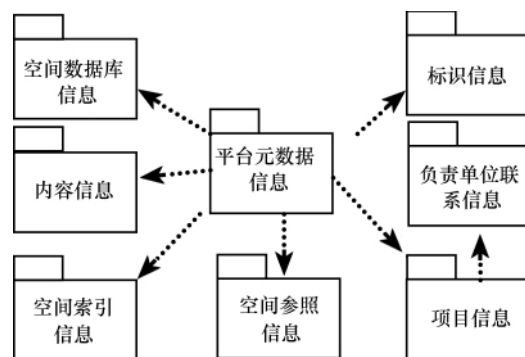


图 1 地质调查平台元数据概念结构

Fig. 1 Conceptual structure of geological survey information platform

在图 1 中, 将地质调查信息平台各分布式空间数据库的信息, 如连接属性等信息定义到平台元数据中, 平台元数据包括空间数据库级和空间索引图层级两级元数据结构, 为空间索引体系的建立提供了良好的数据基础。

3 地质调查信息平台空间索引优化设计

在地质调查信息平台中, 根据当前空间数据生产部门的实际情况, 如果把所有空间数据全部部署在地质调查平台运行系统的服务器中进行空间索引等操作是不现实的, 因此平台的空间索引系统采用分布式网络空间数据库结构, 并将平台两级元数据信息同空间索引的基本信息, 如索引图层名称、索引图层最小坐标范围、Hilbert 空间排列编码等数据一起存放在地质调查平台空间索引数据库里。这样, 地质调查信息平台通过对分布式空间数据库和空间索引元数据进行统一定义, 使相关数据使用部门能快速地将目标数据在海量的空间数据中检索出来。

3.1 空间索引体系结构

按照地质调查信息化过程中建立的地理空间数据仓库

数据格式及存储方案的具体要求,将已建立的数据库纳入地质调查信息平台体系结构中。地质调查平台中空间数据的组织采用 ESRI 的地理数据库 (Geodatabase) 数据模型。地理数据库模型通过 ArcSDE 应用程序服务器在标准的关系数据库——Oracle 中实现。ArcSDE 为用户定义了一个开放的数据库接口,实现地质调查平台系统在不同空间数据库对地理数据的管理。同时,地质调查信息平台是基于 B/S 的体系结构,数据的类型和用途不同,数据的共享程度不同,为了提高数据的访问速度,减少用户响应时间达到网络的负载平衡,空间数据系统采用 ArcSDE 支持的分布式网络数据库结构,实现数据的集中管理和应用的目的^[3]。地质调查信息平台空间索引体系结构如图 2 所示。

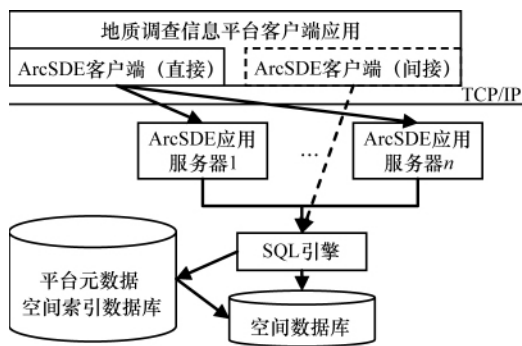


图 2 地质调查信息平台空间索引体系结构

Fig. 2 Geospatial index architecture of geological survey information platform

3.2 空间索引方法分析

地质调查信息平台空间索引的目的是,通过对部署在地质调查信息平台运行系统服务器中的空间索引基础图层(如标准图幅、行政区划图层、成矿区带图层等)进行空间索引定位,将分布在地质调查信息平台外部,并且能与空间索引系统连接成功的空间数据中与索引区域有拓扑关联的图层准确、快速地索引出来。由于空间索引目标图层部署在不同系统的分布式空间数据库结构中,为了减轻地质调查平台系统运行负担,本地服务器只存储了目标图层的相关空间索引信息,而 GIS 软件中进行相关空间分析所需要的大量的空间数据没有存储在本地服务器,因此直接应用 GIS 软件进行分布式空间数据分析在地质调查信息平台空间索引系统中是不现实的。在系统中,建立索引区域的空间信息与分布式空间数据库中的空间信息像 GIS 软件进行空间分析一样的关联,成为地质调查信息平台空间索引系统的关键。

地质调查信息平台通过在空间索引的二维空间区域填充 Hilbert 空间曲线,并且将曲线划分的网格进行有序的 Hilbert 空间排列编码(图 3)。此时 Hilbert 排列编码与本地服务器中的空间索引基础图层要素形成多对一的映射关系(如图 3 所示,面要素 A 对应 Hilbert 码 1,2,3,4,8,9,14,15;点要素 B 对应 Hilbert 码 13),将该映射关系中所有 Hilbert 排列

编码对应的网格范围信息提取出来与索引目标图层信息中的最小坐标范围进行相关拓扑关系分析,如果符合空间索引系统预设的拓扑关系,则将索引目标图层从分布式空间数据库中提取出来,方便用户对地理空间数据的再利用。

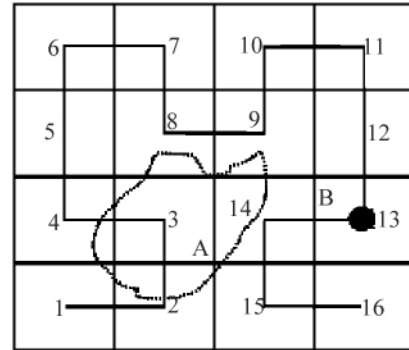


图 3 Hilbert 排列编码与索引要素映射关系

Fig. 3 Mapping between Hilbert ordering code and index features

3.2.1 Hilbert 空间填充曲线分析

Hilbert 曲线源于经典的 Peano 曲线族,是 FASS 曲线的一种^[5]。Peano 曲线族是闭区间单元 $I=[0,1]$ 到闭矩形单元 $S=[0,1]^2$ 的连续映射,也是所有能够填满二维或更高维空间的连续分形曲线的总称,故又称为空间填充曲线。在地质调查信息平台的空间索引系统中,只研究二维 Hilbert 空间填充曲线。

Hilbert 空间填充曲线的描述如下:将某一正方形区域四等分,连接每个小正方形的中心,得到图 4(a)所示曲线即为一阶 Hilbert 曲线;按照一阶 Hilbert 曲线的生成方法进行 4 次迭代,将一阶 Hilbert 曲线在二阶 Hilbert 曲线生成的 4 个象限内做有规则的曲线排列:第一、第二象限曲线为原一阶 Hilbert 曲线原型,第三、第四象限曲线为原一阶 Hilbert 曲线分别顺时针和逆时针旋转 90° ,如此将各个小正方形再细分为 4 个相同的小正方形,并连接各小正方形中心,按照此迭代方法不断细分下去,并按一定规则一一连接,就可得到图 4(b)、(c)二阶、三阶的 Hilbert 曲线。Hilbert 填充曲线,是一种比较好的二维到一维的映射排序方法,实现了网格数组坐标到一维排序序号的映射 $n=f(x,y)$,通过这个映射函数,就可以确定网格单元在一维排序中的位置,找到空间映射关系,为空间索引准备必要的条件。

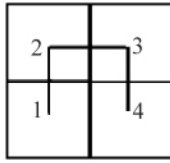
3.2.2 Hilbert 空间排列编码实现

地质调查信息平台空间索引系统中,Hilbert 空间填充曲线对应网格编码,即 Hilbert 空间排列编码的实现,采用基于矩阵运算的 Hilbert 扫描矩阵的生成算法^[6]。

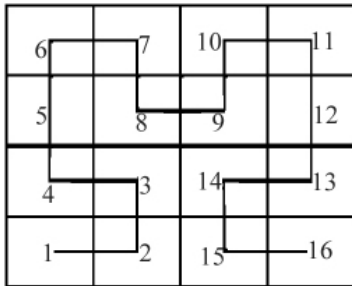
文献[6]详细证明了该算法,具体描述如下。设

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix}_{n \times m}$$

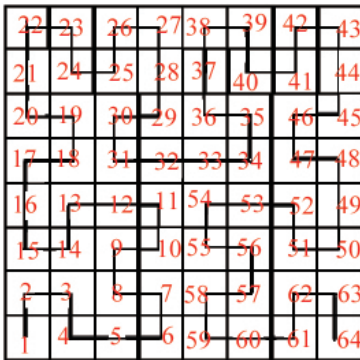
$$\tilde{A} = \begin{bmatrix} a_{1,m} & \cdots & a_{1,2} & a_{1,1} \\ a_{2,m} & \cdots & a_{2,2} & a_{2,1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,m} & \cdots & a_{n,2} & a_{n,1} \end{bmatrix}_{n \times m}$$



(a) 一阶 Hilbert 曲线及其排列编码
(a) The first-order Hilbert curve and ordering code



(b) 二阶 Hilbert 曲线及其排列编码
(b) The second-order Hilbert curve and ordering code



(c) 三阶 Hilbert 曲线及其排列编码
(c) The third-order Hilbert curve and ordering code

图 4 Hilbert 曲线及其网格分割示意

Fig. 4 Mesh segmentation for the Hilbert curve

$$\hat{A} = \begin{bmatrix} a_{n,1} & a_{n,2} & \cdots & a_{n,m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ a_{1,1} & a_{1,2} & \cdots & a_{1,m} \end{bmatrix}_{n \times m} \quad (1)$$

记 2^n 阶 Hilbert 曲线扫描矩阵为 H_{2^n} , 且当 $n=1$ 时, 即 $H_2=$

$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$, 则构造 Hilbert 曲线扫描矩阵的递推算算法为

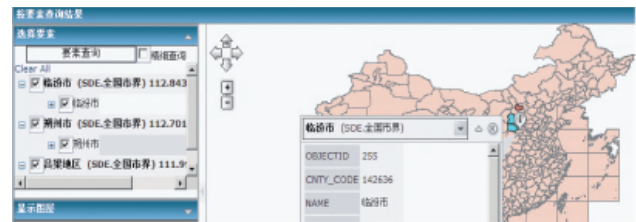
$$H_{2^{k+1}} = \begin{cases} \begin{bmatrix} H_{2^k} & 2^{2k}E_{2^k} + (H_{2^k})^T \\ (2^{2(k+1)}+1)E_{2^k} - H_{2^k} & (3 \times 2^{2k}+1)E_{2^k} - (\tilde{H}_{2^k})^T \end{bmatrix} & k \text{ 为偶数} \\ \begin{bmatrix} H_{2^k} & (2^{2(k+1)}+1)E_{2^k} - \tilde{H}_{2^k} \\ 2^{2k}E_{2^k} + (H_{2^k})^T & (3 \times 2^{2k}+1)E_{2^k} - (H_{2^k})^T \end{bmatrix} & k \text{ 为奇数} \end{cases} \quad (2)$$

式中 E_{2^k} 为相应阶数的单位矩阵。

4 空间索引方法的应用

通过对平台元数据的构建、空间索引数据库的设计、空间索引算法的优化,地质调查信息平台的空间索引体系已应用在资源与环境遥感数据服务系统中^[4,7-8]。数据服务系统提供对数据仓库数据集合的访问,访问的资源通常通过一个名称如标识符等被引用,然后数据服务系统找到这个资源,系统通过生成指定阶数的 Hilbert 空间排列编码(一般不超过 9 阶),遍历空间索引数据库中分布式空间数据的最小坐标范围,与 Hilbert 空间排列编码对应的每一网格坐标值进行空间比较,如果符合条件即为索引目标图层,实现了空间索引的目的。空间索引的应用实现如图 5(a)所示。

在索引体系中,在索引基础图层上选中空间索引的目标区域,显示将要进行索引的基础图层的空间要素和范围,确定索引范围后,点击“按与要素交集查询”,系统运行空间索引算法生成 Hilbert 排列编码与索引要素范围映射表,遍历空间索引数据库中所有空间坐标数据,查找出符合条件的图层信息,其结果如图 5(b)所示。经验证和测试,该空间索引体系的空间索引效率和索引精确度均在之前设计要求范围之内。



(a) 确定空间索引范围

(a) Delimiting spatial index range



(b) 空间索引结果

(b) Result of spatial index

图 5 空间索引实现实例

Fig. 5 Application of the spatial index

5 结论

地质调查信息平台的建设旨在根据《地理信息元数据》建立地质调查信息平台元数据结构,空间数据生产者需要对存入空间数据库的空间数据按平台元数据标准进行空间元数据描述,对于平台索引系统的用户也就是空间数据的消费者,不需要将所有空间数据库中图层等空间数据存储到索引系统运行的服务器中,只需将空间数据的相关元数据信息和空间索引信息导入空间索引系统服务器的空间索引数据库

中,即可进行分布式空间数据库的空间索引。与 Web GIS 开发平台下的空间索引速度取决于空间分析函数的运算速度和运算量不同的是,取决于本空间索引体系索引速度的关键是生成 Hilbert 空间填充曲线的阶数问题,由于 Hilbert 阶数是系统并可以控制的,因此本空间索引体系索引速度是可以控制的。运行空间索引时,平台通过生成 Hilbert 空间排列编码,将空间索引数据与索引条件之间建立一维的映射关系表,以记录的形式存储在关系型数据库中,这样当空间索引系统处理记录时就大大提高了空间索引的效率,从而促进了地理空间数据的再利用,完善了数据共享和信息服务体系建设最终目标。但由于本索引体系中 Hilbert 填充曲线是基于矩形二维填充曲线,因此若 Hilbert 阶数过低,当空间填充曲线与空间索引图层建立映射关系时,会产生一些冗余数据,虽然索引效率很高,但会影响空间索引精度。因此用户需根据实际需求权衡空间索引效率和空间索引精度之间的关系。

通过对资源与环境遥感数据服务系统的设计与研究,发现基于《地理信息元数据》的空间索引体系的设计,在地质调查信息平台建设中空间数据集成共享和分布式空间索引方面有良好的应用效果,并期望在面向 Web 的其他分布式空间数据集成平台中得到广泛的应用,以建立一个标准的空间数据服务的基本环境框架。

参考文献 (References)

- [1] 韩新春. 空间元数据和数据集一体化管理模型及应用研究 [D]. 成都: 成都理工大学信息管理学院应用数学系, 2005: 10-15.
Han Xinchun. Application research of management model based on spatial metadata and dataset [D]. Chengdu: College of Information Management and Applied Mathematics, Chengdu University of Technology, 2005: 10-15.
- [2] Xue Tao, Diao Mingguang. Unified platform about information issuing, integration and management of meta-database and spatial-database based on geology metadata standatd [J]. *Earth Science Frontiers*, 2009, 16(77): 288.
- [3] 中华人民共和国国家质量监督检验检疫总局. GB/T 19710—2005 地理信息元数据[S]. 北京: 中国标准出版社, 2005.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. GB/T 19710—2005. Geographic information—metadata [S]. Beijing: China Standards Press, 2005.
- [4] 陈占龙, 吴信才, 谢忠, 等. 分布式空间数据索引机制研究 [J]. *微电子学与计算机*, 2007, 24(10): 54-57.
Chen Zhanlong, Wu Xincan, Xie Zhong, et al. *Microelectronics & Computer*, 2007, 24(10): 54-57.
- [5] 陈宁涛, 王能超, 陈莹. Hilbert 曲线的快速生成算法设计与实现[J]. *小型微型计算机系统*, 2005, 26(10): 1754-1756.
Chen Ningtao, Wang Nengchao, Chen Ying. *Journal of Chinese Computer Systems*, 2005, 26(10): 1754-1756.
- [6] 王笋, 徐小双. Hilbert 曲线扫描矩阵的生成算法及其 Matlab 程序代码 [J]. *中国图象图形学报*, 2006, 11(1): 119-122.
Wang Sun, Xu Xiaoshuang. *Journal of Image and Graphics*, 2006, 11(1): 119-122.
- [7] Diao Mingguang, Xue Tao, Pan Wenyong. Design of universal authority management model of online teaching system [C]//Proceedings of international conference on computer science and software engineering, CSSE 2008. 2008: 610-613.
- [8] 刁明光, 薛涛, 李建存, 等. 资源与环境遥感项目成果信息服务系统权限模型的设计与实现[J]. *国土资源遥感*, 2009, 20(4): 101-104.
Diao Mingguang, Xue Tao, Li Jiancun, et al. *Remote Sensing for Land & Resources*, 2009, 20(4): 101-104.

(责任编辑 岳臣)

学术动态

“中国科协第五届学术交流 理论研讨会”征文

中国科协学会学术部将于 2010 年 7 月 10-11 日在绍兴召开“中国科协第五届学术交流理论研讨会”。会议主题: 学科发展与科技创新。

征文范围: 国内外学科发展与科技创新的关系; 学术建制与科技创新的关系; 学派及其对学科发展、科技进步的影响; 学术交流在学科发展、科技创新中的地位及其作用等。截稿日期: 2010 年 5 月 25 日。

联系方式: 北京北三环西路 32 号恒润中心 1707 室, 中国工程图学学会 (100086), 杨洁; 电话: (010)62165983; 电子信箱: cegs2009@126.com。

中国科协学会学术部学术交流处, 郭晓芹, 赵崇海; 电话: 010-68580342, 010-68524993。



更正

本刊 2010 年第 28 卷第 7 期第 8 页封面图片说明——“RE211 晶须研究”中, 金刚论文的刊登页码和题目有误, 正确的应为“本期第 19~21 页刊登的金刚的文章‘RE211 晶须实验研究进展’, 在这方面做了一些有意义的工作。”特此更正, 并向该文作者和广大读者真诚致歉。

《科技导报》编辑部