

请参阅以下出版物的讨论，统计数据和作者个人资料：<https://www.researchgate.net/publication/280964271>

## 基于Geohash指数的企业空间数据模型

会议论文 · 2015年8月

DOI : 10.1109 / ICEEI2015.7352548

引文

7

5位作者，包含：



万隆技术学院

288 刊物 430 引文

[查看个人资料](#)



迪基·普里玛·萨蒂亚 ( Dicky Prima Satya )

万隆技术学院

5 刊物 14 引文

[查看个人资料](#)

阅读

1,483



多迪佛法

万隆技术学院

10 刊物 15 引文

[查看个人资料](#)

该出版物的一些作者也在从事以下相关项目：



地理空间搜索引擎 [查看专案](#)



特定特征表示和识别策略的人脸识别 [查看项目](#) Iping Supriana

# 基于Geohash指数的企业空间数据模型

Iping Supriana Suwardi, Dody Dharma, Dicky Prima Satya, Dessi Puji Lestari

ITB电气工程与信息学院

印度尼西亚万隆

{iping, dody, dicky, dessipuji} @ informatika.org

**摘要**——由称为地理信息系统 ( Geographical Information System / GIS ) 的应用程序包装和处理的空间数据。在基于桌面的GIS中，仅当将各种基本数据加载到应用程序数据库中时，才进行空间信息服务。信息技术的发展提供了可随时访问的，全球范围的基于Web的地图系统，例如Google Maps。任何用户都可以轻松地将信息内容添加到地图中。到目前为止，谷歌地图提供的基本数据或用户添加的数据都是松散连接的，这意味着数据之间的链接最少。因此，仍然需要为数据或公司之间相互密切联系的公司或政府带来更大利益的数据。由于要管理的空间数据量很大，因此查询性能的可伸缩性将是一个挑战。为了预料到这一点，我们在建议的空间数据模型的基础上描述了一项改进。我们使用了一种特殊的数据，该数据是通过从空间数据的经纬度对获得的位进行交织而得出的，该字符串称为geohash。地理哈希可以用作空间数据表中每个对象的索引。Geohash字符串越长，其引用位置周围的边界框越精确。这种方法将提高公司GIS数据表中单个甚至一个空间数据集合的查询过程的性能。这项研究的主要目的是提供信息服务以及Google Maps拥有的各种基本空间数据的可用性。

不同的公司将有不同的数据需求和不同的数据表示形式。

信息技术的发展提供了可随时访问的，全球范围的基于Web的地图系统，例如Google Maps。任何用户都可以轻松地将信息内容添加到地图中。这些用户添加的信息也可以作为其他任何用户可以读取的信息注册到Google Maps中。因此，可以轻松搜索和了解公众所需的各种重要信息，例如世界各地的医院位置，政府机构的位置，运动场所或通过提供数据而在任何位置。

已经并且将继续使用现代工具或小工具（例如GPS或智能手机），位置感知服务以及众包互联网地理空间信息来收集大量地理数据。地理空间数据的规模通常太大，而且结构过于混乱，没有足够的关系说明。

到目前为止，Google Maps提供的基本数据和用户添加的数据都是松散连接的，这意味着数据之间的链接最少。因此，仍然需要为数据或公司之间相互密切联系的公司或政府带来更大利益的数据。例如，国家电力公司（PLN）需要管理空间数据，例如变电站的位置，电线杆的位置，电线杆之间的电线以及客户的位置。换句话说，诸如Google Maps之类的基于Web的地图系统无法有效地用于分析目的，尤其是对公司分析有利。在上面的示例中，当变电站遭受雷击时，需要进行分析以识别将受到断电影响的任何客户；

**关键词**—geohash，地理信息系统，空间数据模型，制图系统

一世。 一世 简介

现在，现代技术可以改善地理或地理空间数据的获取，分配和利用（Craglia，2006年）。有许多流行的基于Web的地图绘制技术。通过使用该技术，人们可以基于地球坐标获取任何信息。该技术非常适合个人和日常使用，但是需要将其用于公司或政府数据需求。

地理空间信息对于促进经济发展和改善自然资源管理至关重要。将地图技术用于公司或政府的需求将涉及数据的准备以及非常详细的地理数据之间的关系（例如，支持种植园管理，税收，城市布局，道路质量，河流质量的利益）等）。

为了为上述服务模型提供便利，本研究提出。这项研究的主要目的是提供信息服务以及Google Maps拥有的各种基本信息的可用性。本文重点介绍了我们最近在空间数据管理的理论 and 应用研究中所做的努力。我们正在提议一种新的空间数据管理模型，该模型将接受空间数据结构并创建与基本空间数据的关系。

二。最先进的

信息技术的发展使组织能够执行数字地理数据

管理。通过各种行业的地理分析，数字地理数据已广泛用于帮助解决关键问题。除其他外，地理数据的使用可以通过细分客户数据来帮助分析地理数据，以进行营销优化；协助城市管理和运输；自然资源管理；模拟；和环境保护[1]。如今，地理数据的使用已渗透到各种移动设备中，以帮助人们获得与服务最接近的个人信息

通常，地理数据被称为“地理信息系统”（Geographical Information System / GIS）的应用程序包装和处理。GIS具有执行空间分析的能力，这是需要访问对象的属性和位置信息的一组分析方法（Goodchild（1988）在[2]中）。空间数据由与属性表相关的图形地图组成，这使用户能够

快速可视化并执行查询。在实施中，许多GIS具有读取和存储空间数据的不同方式。

在早期阶段，图形信息服务是作为计算机辅助设计应用程序创建的，它提供了在计算机上画线以设计不同的绘图兴趣的服务。图形信息服务的开发是继续特定于映射的，其提供了在计算机上划一条线的服务，以便于在本地计算机（台式机）上运行地图。进一步的发展是不仅可以在桌面平台上运行，而且可以在Web平台上运行，从而再次提供更轻松地创建线和曲率的服务。

Openshaw（1987）评论说，这样的系统基本上只关注于描述地球表面，而不是对其进行分析（Openshaw（1987）在[3]中）。在这样的系统中，仅当各种基本数据已加载到应用程序数据库中时，才会发生空间信息服务。对于企业而言，构建这样的系统是一个真正的困难，因为他们必须提供基本的空间数据，其特定的空间数据并定义关系。

在Google Maps服务中，用户可以直接访问信息，而不必担心输入基本空间数据。已经提供了诸如坐标，道路，边界等基本数据。另外，Google Maps方便了希望根据自己的利益输入数据的用户。系统开发的机会也成为可能，因为Google Maps具有其他应用程序可以访问的基本功能（通过Google Maps API工具）。这样开发人员就可以开发应用程序，这是对Google Maps满足公司需求的各种基本信息的使用的扩展。

GIS社区似乎对两个简单的命题达成了广泛共识：作为一种技术，GIS有潜力支持许多不同类型的分析，而这一潜力尚未实现[5]。

现在，我们可以获得更多样化，动态和详细的数据。一般来说，地理学和空间科学已经从数据匮乏的时代过渡到了数据丰富的时代。广阔而高分辨率的空间和

时空数据为获取新知识和更好地理解复杂的地理现象（例如人与环境的相互作用和社会经济动态）提供了机会，并解决了现实世界中紧迫的问题，例如全球气候变化和大流行性流感[4]。

为了获得这些知识，需要分析空间数据。但是，为了做到这一点，必须以一种结构来呈现数据，并在来自不同地理信息基础结构的数据之间建立清晰的关系。

为了提高地理信息基础设施的效率和互操作性，许多区域和全球计划都致力于建立开放标准和协议。内容通过规范和标准化的服务类型进行管理。这在分布式环境中强加了地理空间内容的独特生命周期，可以在图1所示的四个步骤中进行描述。首先，必须使内容在分布式系统中可用，即，内容必须在发现和访问服务等标准服务中发布。其次，用户需要发现将通过使用这些服务最终访问的内容（第三步）。最后，用户处理内容并生成新内容，应将其集成并在分布式系统关闭周期中发布[5]。

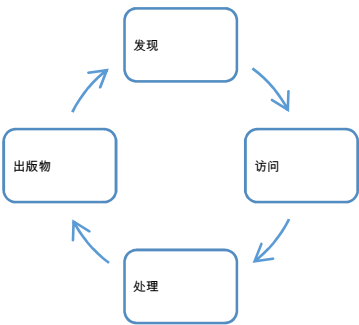


图1. 地理信息基础架构中的内容生命周期[5]

如今，由于许多地区和全球性的制图举措，因此制定了许多标准，并且并非每个政府或公司都遵守这些标准。由于不同的数据需求，他们通常使用自己的标准开发内部应用程序。

三，这项研究

图形和非图形数据管理的概念是在计算机图形学和人工智能ITB实验室（NGGIS（下一代图形信息系统）[6]）中开发的，它提供了可以使用数据的基于桌面的GIS系统的服务。在网络中。

这项研究旨在为基于Google Maps平台的公司模型开发空间数据管理技术。这项研究的结果可以用作：

1. 直观地管理公司数据的基础-根据Google地图

该模型可以作为各种公司（尤其是印度尼西亚）的数据管理和一般准则的基础，从而基于Google Maps可视化其数据，这对印度尼西亚的业务发展很有用

面对全球竞争，这扩大了印度的经济发展。相反过来会加速麻醉。离子系统

2. 企业信息之间的联络  
信息可视化。借助由此产生的技术，可以更直观地呈现，并且可以在Google Maps上直观地呈现信息层。信息将是借助这种可视信息，可以实现公司利益更加有吸引力。nd被定位为蟑螂，分析认为

基于研究的目的和先前的研究，我们修改了pr等人的模型和提出的结果。期。(2013)解释NGGIS数据生命周期。

首先，必须将AV制成地理空间数据，但必须以易于于NGGIS，mple标准公司遵守的SIM卡形式发布内容。NGGIS w构建并定义系统标准将学习数据数据上已经存在的数据之间的关系。以及其他

其次，用户需要发现做空间的内容桥接应用程序。其通过使用广告文字查询。n，由于用户的简单性，最终选择了显示为流行的基于Web信息层技术。y和熟悉并产地图技术之上的查询结果。最后，用户处理内容的内容，全新的出版在机械上。管该使用与我相同的标准进行整合和关闭管理公司的空间毫秒，以gy为基础

期望该系统能够使用流行的基于Web的地图绘制技术满足各种形式通常需要的数据需求，地理信息的表示

我们的研究将集中在三个定义数据模型上；(2)定义主要方面：(1)tion搜索文本查询和空间查询的信息；模拟分析。模型和(3)模型，以及

本文必须解释的第一个研究问题是定义足够简单的空间解决了将是研究进行的数来遵守。从s开始，数据模型有一些需求：数据模型，

(1) 数据模型必须灵活 解决数据实体。改变的动力  
不同企业的要求(2) 数据模型必须适应

需求/数据结构。

IV. HE ROPOSED D ATA M我们成功DEL国家模型。基于表的立了基于相关关系模型的数据模型，该模型成功地将预定义e数学。基准数据库，与有用的表示法相结合。但在我们特定于

空间信息表示空间数据表。每个表都分配有一个un作为unique名称的集合。一中的un表示集合标题之间的关系，标题包含不同的名称，排序。列ch列有于eac

一组可能的值，如图2所示。ed域。结构是

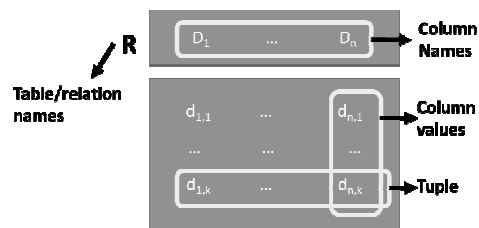


图2.每行的结构是有序的地理空间信息

值的元组 $\langle d_1, d_2, \dots, d_n \rangle$   
这样每个值  $d_j$  在  $t$  的领域 - 第列，用于元组的序列 $\langle d_1, d_2, \dots, d_n \rangle$ ， $j = 1, 2, 3, \dots, n$ 。在数据库中，se必须是不  
不同的，但在我们的da序列中，不一定是数模型，我们提出y与众不同，  
学集合论。这在众多的地理数据中有很多只是 像在  
来源。是因为存在概率ta，数据不是唯一的，并且

关系定义为元素之间的对应关系，在的形式化描述，而表只是e关系本身[7]。  
我们的上下文中，由mea表示的关系的通过观察表格中的ns，我们可以陈述地理空  
能表示

间数据模型：umn相对于

(1) 每个colu的值 (2) 行不是必需

的

彼此 (3) 列的顺序

mn是无关紧要的，因为它们是不同的，而na (4) 标识行m的顺序是按位置的，可能是相关的，因为它们也是按位置的。at包含唯一值，该值可以与行的由内容标识，当有列可以用作键相关。黎凡特每行没有列时。在这种情况下，则相对于包含按e唯一值的按位置称，系统插入行号。当存在相同的表时创建识的关系将创建一个列。该表为我列出了其他的主表

具有唯一值的表之间的表列名称之间的  
的关系将是comcom表。

让  $R_1(YX)$  和  $R_2(XZ)$  两张表由伊利代表  $y$  和  $X$  是B的第n列，因此  
个和 乙连续的A和  $X$  和  $Z$  是列之间的关系  $YZ = X$ 。  
一个和 乙是关系  $X$  包含属性X的A wi值中的  
元组的串联。  
的加入  $R_1$  和  $R_2$  并且是由B中具有相同e  
关系的第1个元组产生的所有元组，如图1所示。

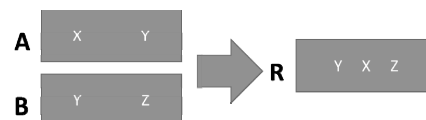


图3.两个spa的关系

具有公共列的空间表

如果一个和乙没有通用列，然后没有笛卡尔积。每它们的关系是与on上的其表都可以通过在表之间创建笛卡尔乘积笛卡尔积来创建ar孩子集的关系。图4。

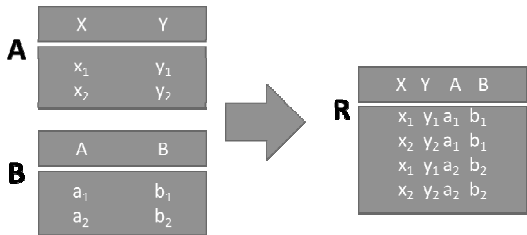


图4。没有c的两个空间表的关系

普通栏

V.G EO H A S H S P A T I A L I N D

敏捷

在空间数据中，基本数据和第三个公共列是三维表示他最有可能二维高度或深处的地球坐标纬度和经度，而不是将坐标数据由nal表示，包括entat存储在数据表中的m中，我们建议存储h坐标数据。

ion。但是，nto的较短哈字符串中的多列会导致Geohash [10]。Base32所提供的基于基础的层次结构中的

为了将坐标数据映射为字符串，我们使用地理编码算法调。geohash可用于对第i个空间数据表进行分区。Geohash提供了位置为表示字符串的地球的grid-b模型。



图5。全球经度的交替划分

-纬度矩形[12] geohash表

全局经度[-180,180]×[-90,90]的交替分隔的二进制字符串中的每个字符 (请参见图5)。f将该矩形分成两个正方形-纬度矩形的第一个除法分([-180,0]×[180]×[-90,90])。点(或更多的基因，在垂直线[-90,90]和[0,呈等距数的左侧以'0'开头，而在地理哈希中的那些以'1'开头。在几何形状)上有一个geo个o中，下一个分割都是水平的；在该点下方的点在'1'。hāsh e，其右半数具有正方形分裂的连续分辨率得以实现。Ge越长，参考周围的边界框e收到一个'0'和es，直越精确。Geohash是通过纬度-经度对的交织得出的；对到所需的eohash字符串为例坐标为-45.995 -41.728，我们可以细分为-d获得更详细的，其位置为d获得的所有级别。Geoh和表示为：

十进制除以空格散列会更长

0010110101011100011此二逆00110001101111000111 d为字母数字

制可以表示为位编码)。每5位为con 08 (32转换为一个字符：00110 00110 111

101 10101 01110 00110，结果为：10 00111

5 p F 6 6 6岁 7

哈希字符串是“5pf666y y7”，代表每个字符40位e位)。er发明了以精度(八个字符，五个c点作为短字符串的geohash

2008年，Gustavo Niemeye专门对网络URL (http://www系统中使用的地理编码)进行了目的编码，于2008年2月26日逆geohash.org/)。他通过发布Wikipedia页公共领域[11]。任何非负长度，通常以面命令进入，以增加地理哈希值或分配纯文本，以base-32作为二进制字符串，地理哈希值只能用于网络，而h为5的倍数。在矩形上的递归四叉树。地质哈希提供

地理哈希隐含地定义了世界范围内的一些经度-纬度[13]。

- 每个geohash代表一个经度纬度矩形。到位于初始e 7内的geohas
- 遏制。添加cha指定较小的rectah.angle末尾  
Ĥ Ć tt ttuv。(数字
- 地质哈希提供了覆盖整个地球的矩形的z阶遍历h分辨率。(图6)表示紧密。  
(“dp”很近，没有通用前缀，
- 地区性。共享前缀“dr”)。但是相  
邻的不像9z和dp(图7)

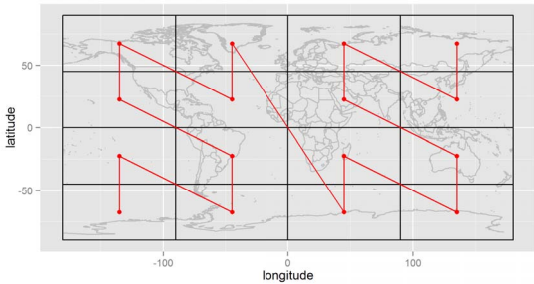


图6.两个spatia的关系对位置进行龔具有共同列[13]的表(纬度，经度)对被忽略码后，将第三个坐标(高度)作为spa龔，因此它不知道真正的ace。对象的3D位置

由于附近的地方经常会出现(而不是在某些极端情况下，赤道或ixix，精度逐渐子午线中)类似的偏好，所以两个地方龔低。共享前缀越长近

e。

为了检索纬度和坐标对表示，应用龔空间的最初ng个目标点开始的d个经度位，算法连续和更多的ely

通过在每个步骤之间交替的枢轴点减少了限制其余地理区域坐标。选择大于f位的s的精确地理区域。如果目标坐标值是轴，则将1位附加到轴；否则，包含0位的集合中。剩下的地理原点然后在nex算法中使用。连续Geohash集会以准确性的x迭代会增加最终的Geohash字符串。一个吸引人的属性是迭代的ac区域共享相似的是，附近的点将生成Geohash字符串。序列越长，两个点越匹配位，近。[9]。

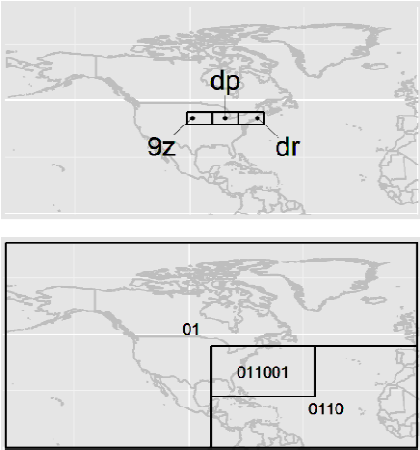


图7。Geohash的位置和Nes 刺

我们可以改进空间da地理编码算法中的索引，因为它使用索引范围的了必须存储在每个inst中的信息。的Geohash可以在不添加位置用于存储数据。间的情况下为空间查询指定搜索空间 前两个会极大地减少标准索引编制。基本数据和第th个坐标，但仅以rge和复数形式存在，即称为区域的特征。通过甚至可以简化商店的收藏

VI. 达美模型国家

如前所述，在空间数据中，最可能的公共列是耳朵，我们知道公共列可以存在n个坐标。空间数据通常是点的较大集合。几个点可以创建线。闭环线可以使用geohash来创建要素以表示案例中的每个点。例如，假设我们要指向表：

{ ( 6 ° 50 ' 53.08 " 小号 107 ° 36 ' 35.32 " E ) , ( 6 ° 51 ' 41.00 " 小号 107 ° 21 ' 22.21 " E ) , ( 6 ° 52 ' 77.32 " 小号 107 ° 24 ' 13.41 " E ) , ( 6 ° 53 ' 53.08 " 小号 107 ° 36 ' 35.32 " E ) } ,

Geohash函数将这些数据压缩为 o :

{ qqqu909r543dz , qqqu2fyym03s , qqqu2u4118hvu , qqqu8b8ppf34c } .

我们可以看到该字符串，因为它们具有通用前缀“ qqqu ”，sa中，因此它们被自动分组，从而节省ane区域（彼此相邻）。地上。这对于在数存储空间和t位置。由于字符串是天真的数据库系统中基于ve数据类型来搜索数据，使用und和过滤时间非常有用 串。

搜寻 算法。

而不是使用桌面g已经创建的Goog地理信息服务，这些基本的有意义的功能Maps（例如地点坐标（点（地区））。（线），道路（线）或城市边界的卫星信息是基本的geo可以创建基于地理空间信息衡以立即使用Basic的公司可以将精力集中信息），它们可特定于地理空间信息Map主要和n上，并将其与Google tion相结合，基本的地理空间信息 如图8所示。

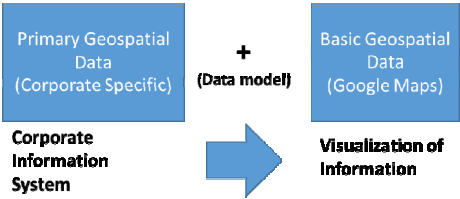


图8.公司GIS为了关联原始始Google Maps Platform上将地理空间数据理空间数据和关联其他主要/次要地理款基本的主要地理空间数据转换为空间数据据模型。

例如，State Electri管理诸如选举的业/ Company（PLN）需要将变电站的位置间数据可以根据其定义位置，如表1所示，设置为Tricity。他们是地球坐标。那个地点

表一 小号 基站 大号 职业		
变电站Latitude	ST001	经度08 " S 107 ° 36 ' 35.32 " E
	6 ° 54 ' 53.0 03 " S 107 ° 25 ' 15.13 " E	
ST002	6 ° 40 ' 44.0	

通过使用geohash，我们可以形成，以更紧凑的方式保存数据如表II所示

表II。 小号 热情 n 大号 职业 w ^ l t h G 环境乙族 H 灰	
变电站G ST001	Geohash
	qqqu909r543dz
ST002	qqqu6h2jxkem7

加载此表后，它会自动与哈希字符串系统基本空间数据输入NGGIS，并将ST相关联，并在Google Maps di顶部创建001和ST002解码为新的显示。由那个变电站个点层 服务

如果我们按照表3所示定义regio ST0 01 3，系统会建立联系

自动根据变电站列在表2和表3之间进行设置，因为它们具有相同的列名。

另一方面，系统还将表2与基本空间数据相关联，并创建一个区域，因为它具有具有相同变电站值的geohash集合。区域是具有特定属性的地球坐标的集合，最好是闭环的。该区域将作为新图层绘制在Google地图显示的顶部。

表III。 小号 热情 [R 变数

变电站Geohash	ST001
	qqu909r543dz
ST001	qqu2fyym03s
ST001	qqu2u4118hvu
ST001	qqu8b8ppf34c
ST002	...
ST002	...

另一个实例展示了使用geohash字符串将位置存储在关系数据库系统中的优势，例如，我们想要在地图上进行放大和缩小并显示网格正方形上有多少个点的摘要，我们可以使用相对于关系查询语法中缩放分辨率的geohash前缀长度：

从位置GROUP BY SUBSTR中选择SUBSTR ( geohash , 0 , 2 ) , COUNT ( \* ) ;

查询执行将产生以下结果：

“aa”，67“
ab”，456...”
qq”，128
... “zy”，8

在流行的RDBMS MYSQL中，我们可以使用geohash函数进行查询，例如：

SELECT ST\_AsText ( ST\_PointFromGeoHash ( @ gh , 0 ) ) ;

该查询将给出结果：

点 ( 45 -20 )

在上面的位置表的geohash列上有一个简单的索引，使我们可以将所有查询集中在地图上的特定区域。

七。结束语

我们的研究集中在三个主要方面：（1）定义数据模型；（2）定义用于文本查询和空间查询的信息搜索模型；（3）模型与仿真分析。通过定义足够简单的空间数据模型已经解决了第一方面。

第二方面是通过使用我们提出的数据模型得到部分回答的，该模型通过geohash字符串进行了增强，以表示空间数据的坐标字段。我们可以利用现有的RDBMS功能，该功能是为支持字符串数据操作（例如搜索或操纵字符串）而开发的。即使在像MySQL这样的流行RDBMS中，也具有特定的Spatial Geohash函数来编码或解码哈希字符串。这一事实将提高企业GIS系统的整体性能，尤其是在数据过滤方面。为了充分回答我们研究的第二方面，我们需要更详细地探索与信息搜索模型有关的信息。

致谢

感谢印度尼西亚共和国教育和文化部的Direktorat Jenderal Pendidikan Tinggi ( DIKTI )，该研究通过Penelitian Prioritas国家总体规划Percepatan和Perluasan Pembangunan Ekonomi Indonesia 2011-2025 ( PENPRINAS MP3EI 2011-

2025 )。

R参考

[1] ESRI, “谁使用GIS?”, [http : //www.esri.com/what-is-gis/who-uses-gis](http://www.esri.com/what-is-gis/who-uses-gis) , 在2013年3月14日访问。[2] AC Gatrell, 空间和地理数据的概念。第119 – 134页，1991年。  
[3] M. Goodchild, R. Haining和S. Wise, “集成GIS和空间数据分析：问题与可能性”，《国际地理信息系统杂志》，第1卷。6号 5, 第407-423页，1992年。[4] D. Guo , J. Mennis , “空间数据挖掘和地理知识发现-简介”，计算机，环境和城市系统，第1卷。33, 第403-308页，2009年。[5] S. Trilles , L. Diaz , J. Gill和J. Huerta, “地理空间数据和元数据的辅助生成和发布”，《国际空间数据基础设施研究杂志》正在审查中，于2013年2月27日提交。[6]  
I. Supriana, PR Aryan, “下一代图形信息系统引擎 ( NGGIS )：旅游业应用案例研究”，TSSA和WSSA国际联合会议，2006年。[7] P. Atzeni和V. de Antonellis, 关系数据库理论，本杰明/卡明斯出版公司，1993年。[8] M Malensek , SPallickara , SPallickara ,

“基于多边形的查询使用分布式哈希表评估地理空间数据”，2013 [9] ZBalkić, DŠoštarić和G Horvat , “ Ge oHash和UUID标识符多代理系统”，2012年[10] <http://geohash.org/site/tips.html>

[11] Geohash. <http://en.wikipedia.org/wiki/Geohash>. [线上; 已访问20-2013年六月]。[12] <http://www.bigfastblog.com/geohash-intro>

[13] Fox , C Eichelberger, J Hughes , “ Skylar时空索引在非关系分布式数据库”