

一种基于 REST 的分布式地理资源聚合系统

张 珊 吴健平 胡志明 张 赐

(华东师范大学地理信息科学教育部重点实验室 上海 200062)

摘 要 Mashup 是一种新型 Web 应用,指多种不同类型资源的组合重用。针对此,提出采用 REST(Representational State Transfer)风格面向服务架构,结合 JavaScript API、AJAX 技术,实现分布式地理资源的 Mashup,并构建了上海市绿地社会服务功能分析原型系统。研究表明,该方法可为用户提供良好的分布式地理资源在线聚合体验,更重要的是将简单地理处理资源组合为资源链完成复杂地理处理功能,节约了开发成本。

关键词 聚合 分布式资源 RESTful Web Service 面向服务架构

A DISTRIBUTED GEOGRAPHIC RESOURCES AGGREGATION SYSTEM BASED ON REST

Zhang Shan Wu Jianping Hu Zhiming Zhang Ci

(Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai 200062, China)

Abstract As a new web application, Mashup refers to the integration and reuse of different types of resources. In light of this, this thesis presents a solution to realise the Mashup of distributed geographic resources which uses the service-oriented architecture of Representational State Transfer (REST) style and in conjunction with the technologies of JavaScript API, Asynchronous JavaScript and XML (AJAX). A detailed research on the development and construction of a prototype of Shanghai Greenland's Social Service Function Analyze System is conducted as well. Research shows that the users can experience online aggregation of distributed geographic resources conveniently. More importantly, the combination of simple geographic process resources into a service chain can achieve complex function, thus saving the develop cost a lot.

Keywords Aggregation Distributed resource RESTful Web service Service-oriented architecture

0 引 言

21 世纪互联网的迅速崛起使得万维网成为高效的全球性信息发布渠道,基于互联网技术的地理信息系统——WebGIS 应运而生。然而,目前的 WebGIS 产品大都采用不同的分布式技术,在不同软硬件环境中独立开发。用户想在这样的系统中浏览或整合其他系统的数据、功能资源非常困难甚至是不可能的。伴随着人们对 GIS 需求的不断提高,在 Web 上共享地理空间数据、地理处理服务已经成为 GIS 发展的必然趋势。

Mashup 的出现为上述问题的解决提供了契机。Mashup,中文常译作聚合、混搭,是一种新型 Web 应用程序,借助 Web2.0 在全球迅速传播。目前为止在计算机领域对 Mashup 还没有一个明确的权威的定义。但普遍认为 Mashup 是一种新型的基于 Web2.0 的数据集应用程序,通过多种渠道将两个或以上的外部数据、应用、Web API、UI 组件堆叠、聚合起来创建全新的服务,实现了资源的组合和重用,关注用户的参与、协作,全面改善用户的客户端体验^[1]。短短几年间,Mashup 已经吸引了众多厂商的关注,Amazon、Google、Yahoo!、Microsoft、Flickr 等国际公司均先后开放了丰富的 API 供用户自由 Mashup。综合国内外研究现状,Mashup 还属于一个新兴的研究领域,许多研究才刚刚起步,热点也多集中在 Map Mashup,对其他资源如地理资源的聚合研究还很少。基于此,本文主要研究分布式地理资源聚合,并

以上海市绿地社会服务功能分析系统为原型开发了相关地理资源聚合系统。

论文通过对 REST 技术及 Mashup 框架的研究,采用 ArcGIS Server 将地理资源发布为 REST 式服务,利用 ArcGIS JavaScript API 聚合服务资源,采用 AJAX 技术实现客户端与服务器端数据的异步传输,最终实现了分布式地理资源的聚合。

1 相关技术

1.1 REST

2000 年,Roy Fielding 首次提到了 REST,表述性状态转移这个概念^[2]。其架构主要包含三个重要元素:资源,资源标识和表示。一切能够被命名的东西都叫作资源,在 GIS 系统中,服务提供商所提供的数据服务、地理处理服务,就是“资源”。每个资源都对应一个唯一的 URI 作为资源的标识。客户端可以通过 HTTP 的基本方法 GPPD(GET、POST、PUT、DELETE)得到资源的各种表示如 HTML 页面形式、XML 或者 JSON 数据格式。相比 SOAP 协议栈,REST 不需要额外的 XML 标记,不需要任何开发包,轻量级和高效率的 HTTP 格式可直接被应用,拥有更好的性能和缓存支持,改善了响应时间,改进了用户体验。此外,

收稿日期:2010-12-22。2008 年上海市科学技术项目(F089903)。张珊,博士,主研领域:WebGIS,数据共享。

REST 网址能够被任何搜索引擎检索,这使得它们更容易被发现^[3]。对于服务提供者而言 REST 式服务能降低创建服务的成本,降低托管和维护支持服务的费用。对于服务使用者而言可以降低学习难度,减少构建系统应用所需要的时间和投入。

万维网上有许许多多的 REST 式服务,例如 Amazon S3 允许在 Amazon 服务器上存储数据,并通过 HTTP 或 BitTorrent 来访问那些数据;OpenStreetMap 项目建立了一个可以免费访问的地图数据集,并提供了一个 REST 式接口来访问公路地图,它的资源既不是图片也不是地点,而是构成地图的原始点与向量。在每一个 REST 式服务的背后,都有一个可寻址的、无状态的“超文本链接目标”。

随着 REST 式服务的出现和推广,面向 Web 服务的地理信息共享进入了全新的模式,它以更贴近 WWW 基础协议的方式来实现 Web 服务,大大简化了 Web 服务的设计。

1.2 资源 Mashup

随着越来越多的信息提供者公开自己的 API,各种新型的 Mashup 应用开始在网上出现。一个典型的 Mashup 应用框架由三部分构成:资源的提供者、Mashup 站点和客户端^[4]。资源提供者将资源通过 Web 服务的形式发布。Mashup 站点通常负责使用服务器与客户端的逻辑组合来实现服务的集成。客户端则以图形化方式呈现 Mashup 应用并负责与用户交互。

Mashup 应用构建主要基于 AJAX、Web 服务访问接口、屏幕抓取和语义网这几项技术^[5]。由于本文以上海市绿地社会服务功能分析系统为原型,资源提供者明确,因此本文主要研究 AJAX 和 Web 服务访问接口技术,对于屏幕抓取和语义网技术暂不讨论。目前主流的 Web 服务主要有两种类型:基于 SOAP 协议栈的 Web 服务和基于 REST 风格的 Web 服务。SOAP 协议栈虽然功能强大,但基于 SOAP 的服务是完全预定义的,且协议复杂,不容易被很多开发者使用。而 REST 式的服务,输入网址即可请求服务,使用极其简单,只要求很少的编码工作量,可以减少很多不必要的工作。因此,采用 REST 式 Web 服务作为 Mashup 系统的资源提供者成为首选。客户端的构建则主要采用 Ajax 技术。Ajax 实际上是 XML、JavaScript 等多种技术的组合,通过构建部署于客户端的引擎,使用 XMLHttpRequest 实现客户端与服务器之间的异步查询检索,采用 XML 进行数据交换,使用 JavaScript 和 DOM 进行动态交互与显示。

篇幅所限,本文仅选取上海市绿地社会服务功能分析系统中的“单位面积绿地服务人口数”指标为例说明如何构建基于 REST 的分布式地理资源聚合系统。

2 分布式地理资源

2.1 地理数据资源

系统研究所采用的数据包括 2009 年采集的上海市主要公园绿地的分布、2006 年上海市土地利用类型、上海市街道分布图、上海市行政区划图各街道人口数据(2010 年)、上海市环线(内环线、中环线、外环线)分布地图。首先利用 ArcMap 将上述数据分别制作为 MXD 地图文档,再将地图文档利用 ArcGIS Toolbox 或者 ArcGIS Server Manager 分别发布为 Map Service 作为聚合所需的数据资源。

2.2 地理处理资源

明确“单位面积绿地服务人口数”指标的计算都需要用到

哪些地理处理操作是构建地理处理资源的第一步。分析可知,获取该指标值首先需要生成居住区单元并统计每个街道(镇)居住区单元总数,将对应街道的人口总数平均分配到每个居住区单元。然后获取每块绿地的服务范围,以面积大于 3000m²的绿地为源地做缓冲区。将第一步得到的居住区单元与第二步获得的绿地服务范围进行叠置识别分析,生成居住单元叠置图层。利用拓扑分析的方法,判断居住区单元是否位于绿地的服务范围内。若位于服务范围内,则将居住区单元的编号与此绿地服务范围的编号相关联。最后统计与某个公园关联的居住区单元内的人口总数。这些人口总数被考虑为经常到此绿地游憩的人数,即此绿地服务的总人口数。根据此绿地的面积,即可计算绿地单位面积所服务的人数。模型主要用到 ArcGIS Toolbox 中的 Identity、Buffer、Dissolve、Copy Feature、Intersect、Join Field、Add Field、Calculate Field 等工具。将上述流程拆分为三个子流程并分别建立模型,模型示意图如图 1 所示。

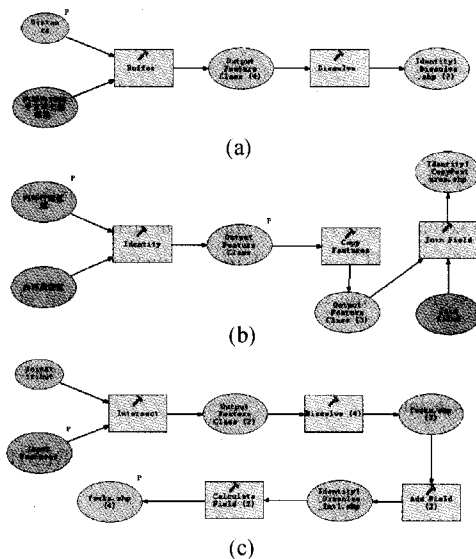


图1 单位面积绿地服务人口数模型

利用 ArcGIS Server 将图 1 中(a)、(b)、(c)三个模型分别发布为 Geoprocessing Services。

3 系统设计及应用

3.1 系统总体架构

根据上海市绿地社会服务功能分析系统的功能要求及需求特点,同时考虑到系统的可扩展性、灵活性,以及和现有系统的风格匹配。系统主要采用 ExtJS 风格。主页面顶部为工具栏,包括常规功能按钮如放大、缩小、漫游、全图、前一视图、后一视图、查询,以及三个特定功能按钮,分别为“地理数据聚合”按钮、“绿地服务率”按钮和“绿地服务人口数”按钮。工具栏下方左侧区域顶部是图层列表,用户可以通过勾选图层复选框来指定图层的显示与隐藏。图层列表下方提供了快速缩放到当前选中图层的“Zoom To Layer”按钮以及设置当前图层的透明度的滑动条。页面中央为地图绘制区,显示当前聚合的地图服务以及地理处理功能执行的返回结果。系统总体架构如图 2 所示。

客户端通过浏览器向 Mashup 站点发出 HTTP 请求,Mashup 站点根据请求内容分别向 REST 式地理数据资源提供者、REST

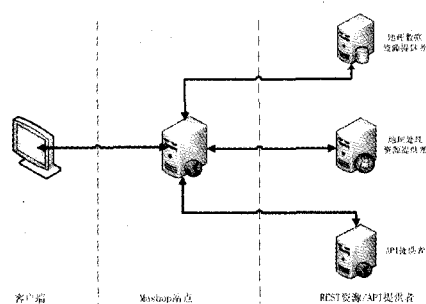


图2 聚合系统总体架构图

式地理处理资源提供者和 API 提供者发送相应请求参数。REST 资源/API 提供者接收请求后分别返回对应结果给 Mashup 站点, Mashup 站点将接收到的 REST 资源聚合后将最终结果返回给客户端。

3.2 应用实例

(1) 地理数据资源聚合

本系统聚合了来自微软的 Bing map, 来自开放街道地图的 OpenStreetMap, 以及来自实验服务器发布的 2.1 节所述的相关数据资源。示例代码及说明如下:

//创建 Bing map 图层

```
veTileLayer = new esri. virtualearth. VETiledLayer ({bingMapsKey: 'Av1bH4keF8rXBtxWOegklgWGCYYz8UGYvBhsWKuvCT76xVFOERk8jkKEDvT'});
```

osmLayer = new esri. layers. OpenStreetMapLayer();

//通过资源地址调用 REST 数据服务并创建图层

```
layer = new esri. layers. ArcGISDynamicMapServiceLayer ("http:// 219. 228. 148. 220. 8399/arcgis/rest/services/greenland/MapServer");
```

//初始化地图

```
map = new esri. Map("center", {extent: customExtentAndSR});
```

//添加图层

```
map. addLayer(osmLayer);
```

```
map. addLayer(veTileLayer);
```

```
map. addLayer(layer);
```

分布式地理数据资源聚合效果如图 3 所示。

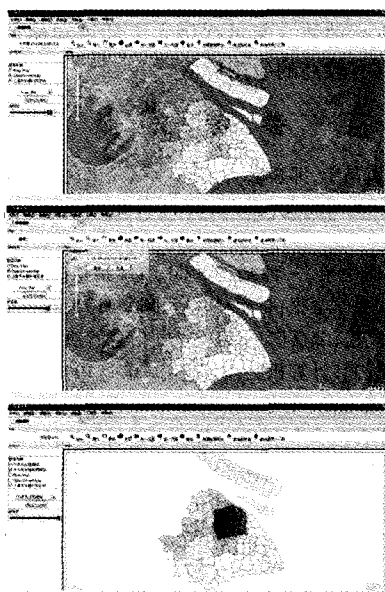


图3 地理数据资源 Mashup

其中,上图为 OpenStreetmap、Bing map 和上海市乡镇行政区划地图服务三个分布式地理数据资源 Mashup 的效果图。单击系

统工具栏上“地理数据聚合”按钮,在弹出对话框中输入添加的地图服务 URL 地址,如中图所示,单击“确定”按钮。下图即为系统在上图基础上,添加的“外环内公园绿地”和“外环内土地利用类型”地图服务,并隐藏 Bing map 以及 OpenStreetmap 资源的效果图。实验表明,采用 REST 式的资源服务,可以很方便地实现分布式地理数据资源的聚合。

(2) 地理处理资源聚合

如果地理处理服务被发布为异步提交模式,用户提交任务后可继续进行其他操作,服务执行完后根据任务唯一的 JobId 属性到服务器指定路径读取执行结果。如果地理处理服务被发布为同步执行模式,则服务执行后客户端可以直接进行可视化显示。在 2.2 节中我们设计了三个地理处理模型分别用于完成绿地缓冲区计算,行政区划与居住区图层的识别以及图层求交和相关字段值计算。通过 ArcGIS Server 将这三个模型分别发布名为 BufDis、JoinCalcu 和 InterCalcu 的 REST 式地理处理服务。利用 ArcGIS JavaScript API,将这三个 REST 式地理处理服务聚合顺次调用,即可实现单位面积绿地服务人口数的计算。服务间输入输出参数的传递主要依赖于 geoprocessor 对象调用 getResultData(jobInfo.jobId, "output", getData) 方法实现。

主要代码如下所示:

```
var gp1 = new esri. tasks. Geoprocessor ("http://219. 228. 148. 220: 8399/arcgis/rest/services/cn/GPServer/BufDis");
```

```
var gp2 = new esri. tasks. Geoprocessor ("http://192. 168. 18. 9:8399/arcgis/rest/services/cn/GPServer/JoinCalcu");
```

```
var gp3 = new esri. tasks. Geoprocessor ("http://219. 228. 148. 114: 8399/arcgis/rest/services/cnexecute/GPServer/InterCalcu");
```

//提交任务

```
gp1. submitJob(params1, gp1.completecallback, errback1);
```

//获取处理结果

```
gp1. getResultData(jobInfo.jobId, "output1", getData1);
```

```
gp2. submitJob(params2, gp2.completecallback, errback2);
```

```
gp2. getResultData(jobInfo.jobId, "output2", getData2);
```

```
gp3. execute(params3, gp3.displayResult2, errback3);
```

通过上述流程,即可实现 2.2 节中三个分布式地理处理资源的聚合。在用户设置不同的服务半径时,系统的计算结果如图 4 所示。

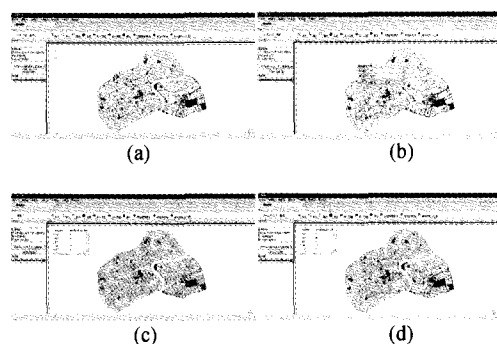


图4 单位面积绿地服务人口数

图3中(a)为基础底图展示,包括上海市内环内3000平方米公园绿地地图服务、内环居住区地图服务和内环行政区划地图服务。单击工具栏上“绿地服务人口”按钮,在弹出对话框中输入绿地服务缓冲半径并选择半径单位,如(b)所示。(c)、(d)分别为缓冲半径500米和100米的成果图,用户还可通过查询以表格的形式查看处理结果。结果表明不同区域的单位面积公园绿地服务的人口数分布并不均匀,越靠近城市中心的区域,公

园绿地的服务压力越大,绿地单位面积服务人口越多,公园绿地服务游人的相对舒适度相对越低。实验表明,通过该方案可以实现分布式地理处理资源的聚合,一方面用户能通过网络共享分布式地理处理资源,另一方面简单功能组合完成复杂任务,为相关部门决策提供了依据。

4 结 语

本文利用 ArcGIS Server REST Service 良好的灵活性和易用性,采用 AJAX 和 ArcGIS JavaScript API 技术,基于 REST 风格,快速构建了上海市绿地社会服务功能分析系统,最终实现了分布式地理资源的聚合。实践证明,该方法不仅可以为用户提供良好的分布式资源在线聚合体验。更重要的是简单的地理处理资源组合为资源链可以完成复杂地理处理功能,节约了开发成本。同时,系统轻量级的客户端使得用户通过浏览器即可查看

分布式地理数据资源、运行分布式地理处理资源,具有重要应用价值,为 GIS 推广起到了积极作用。

参 考 文 献

- [1] Yu Jin, Boualem B, Fabio C, et al. Understanding Mashup Development [J]. IEEE Internet Computing, 2008, 12(5): 44-52.
- [2] Fielding, R. T. Architecture Styles and the Design of Network-based Software Architecture [D]. IRVINE: University of California, Irvin, 2000.
- [3] Leonard Richardson, Sam Ruby. RESTful Web Services [M]. 北京: 电子工业出版社, 2008.
- [4] Merrill D. Meshups: The new breed of Web app [EB/OL]. (2006-08-08) [2010-09-25]. <http://www.ibm.com/developerworks/xml/library/x-mashups.html>.
- [5] 祁晓晖, 郑滔. Mashup 研究综述 [J]. 计算机应用, 2010, 6(30): 203-209.

(上接第 58 页)

际共产主义运动中长期没有解决的一个重大理论问题。

首先将三个例句在北大计算机语言研究所的分词测试平台上^[8]测试,结果如图 1 所示。

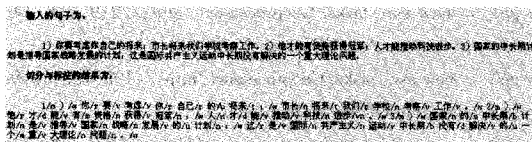


图 1

然后再在猎兔分词平台^[9]测试结果如图 2 所示。

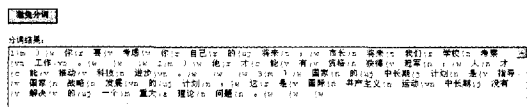


图 2

从两个测试平台来看,三例句中的组合型歧义字段“将来”、“才能”、“中长期”都未能准确消除。因为三句子中的后半句都错误的采取从合切分。

依据我们设计的算法人工计算各结果词的支持度如表 1 所示。

表 1 歧义字段可能切分结果词与前后搭配词的支持度表(1,2 表示出现顺序)

| 组合型歧义字段 | 才能 1 | 才能 2 | 将来 1 | 将来 2 | 中长期 1 | 中长期 2 |
|------------|------|------|------|------|-------|-------|
| 从合前共现 | 1 | 0 | 1 | 0 | 1 | 0 |
| 从合后共现 | 1 | 1 | 0 | 0 | 1 | 1 |
| 从分前共现 | 0 | 1 | 0 | 0 | 0 | 1 |
| 从分后共现 | 1 | 1 | 0 | 1 | 1 | 1 |
| Support(合) | 2 | 1 | 1 | 0 | 2 | 1 |
| Support(分) | 1 | 2 | 0 | 1 | 1 | 2 |

从表中可以看出:三例句中的歧义字段“将来”、“才能”、“中长期”前半句支持从合,后半句支持从分,符合切分预期。

3.2 实验模拟与评估

为方便演示我们所设计的算法,我们选取 8 个组合型歧义字段,从语料库中下载包含歧义字段的一定数量的语料,利用分词字典与人工相结合的方式鉴别前后搭配所成词的合理性及其

切分的正确与否。实验结果如表 2 所示。

表 2 实验结果

| 歧义字段 | 例句数 | 正确率 | 歧义字段 | 例句数 | 正确率 |
|------|-----|-------|------|-----|-------|
| 才能 | 86 | 97.6% | 中学 | 125 | 98.4% |
| 将来 | 98 | 96.9% | 研究所 | 135 | 100% |
| 中长期 | 78 | 98.7% | 及其 | 68 | 97.1% |
| 上来 | 96 | 97.9% | 一生 | 85 | 97.6% |

结论:我们的方法具有较高的可信度。从实验结果来看,利用分词字典,从共现支持度的角度消除组合型歧义字段是可行的。

4 结 语

本文讨论了基于关联规则的组合型歧义的消解策略,方法的最大特点是从可能的切分方式所得结果词与前后搭配所构成的新词的支持度出发,构造从分或从合切分支持度因子,对组合型歧义字段进行消歧。通过样例人工计算和模拟实验验证,该方法可行并具有较高的应用价值。

参 考 文 献

- [1] 孙茂松,黄昌宁,邹嘉彦. 利用汉字二元语法关系解决汉语自动分词中的交集型歧义 [J]. 计算机研究与发展, 1997, 34(5): 332-339.
- [2] 刘禹孜,何中市. 一种基于 SVM 和规则消除组合型歧义的算法 [J]. 重庆大学学报:自然科学版, 2005(10): 50-53.
- [3] 肖云,孙茂松,邹嘉彦. 利用上下文信息解决汉语自动分词中的组合型歧义 [J]. 计算机工程与应用, 2001(19): 87-90.
- [4] 丁德鑫,曲维光,徐涛,等. 基于 CRF 模型的组合型歧义消解研究 [J]. 南京师范大学学报:工程技术版, 2008, 8(4): 73-77.
- [5] 冯素琴,陈惠明. 基于语境信息的汉语组合型歧义消歧方法 [J]. 中文信息学报, 2007, 21(6): 13-17.
- [6] 曲维光,吉根林,穗志方,等. 基于语境信息的组合型分词歧义消解方法 [J]. 计算机工程, 2006, 32(17): 74-77.
- [7] 秦颖,王小捷,张素香. 汉语分词中组合歧义字段的研究 [J]. 中文信息学报, 2007, 21(1): 3-8.
- [8] <http://www.icl.pku.edu.cn/>.
- [9] <http://www.lietu.com/SCSeg.jsp>.