

# 从万维网中提取模式和关系

谢尔盖 • 布林计算机

科学系

斯坦福大学 Sergey

CS 斯坦福教育

摘要万维网是巨大的信息资源，同时它的分布非常分散。特定类型的数据（例如餐厅清单）可能散布在成千上万种不同格式的独立信息源中。在本文中，我们考虑了以下问题：从所有这些来源中自动提取此类数据类型的关系我们提出了一种利用模式和关系集之间的对偶关系从小样本开始发展目标关系的技术。要测试我们的技术，我们使用它来提取关系万维网上的作者标题对

## 介绍

万维网可以提供几乎所有类型的信息，从 DNA 数据库到履历再到喜欢的餐馆列表，都是如此。但是，如果可以从中提取这些信息，这些信息通常会散布在许多使用不同格式的 Web 服务器和主机中互联网并将其整合为结构化形式，它们将形成前所未有的信息来源。它将包括最大的国际人员目录，最大，最多样化的产品数据库，最大的学术著作目录和许多其他有用资源

使用特殊编码的包装器或过滤器 Tsi MOS 集成大量信息源已有大量工作，但是创建这些信息源可能很耗时，通常用于数以千计的源。在本文中，我们解决了从中提取关系的问题。可能在万维网上保持这种关系的成千上万个源我们的目标是发现信息源，并完全自动地或以最少的人工干预从中提取相关信息。

在本文中，我们考虑提取书本关系的问题

Web 上的作者标题对直观上，我们的解决方案如下工作：首先，在测试中使用少量的作者标题对种子集，我们使用一组

第五本书，然后我们发现这些书在网络上的所有出现次数，从这些出现中我们识别出书的引用模式，然后我们在网络上搜索这些模式并找到新书，然后我们可以将这些书和

找出它们的所有出现，并从中产生更多的模式。我们可以使用这些新模式来找到更多的书，依此类推，最终，我们将获得大量的书和模式来查找它们。

## 模式和关系的对偶

我们提出的方法称为 DIPRE 双重迭代模式关系扩展。它依赖于模式和关系之间的对偶关系，我们在下面进行解释

### 问题

在这里，我们更正式地定义我们的问题

令  $D$  为非结构化信息的大型数据库，例如万维网，令  $R$  为目标关系。 $R$  的每个元组  $t$  在  $D$  中出现一次或多次。每次这样的出现都由  $t$  的所有字段组成，表示为字符串在 Web 的情况下，在  $D$  中彼此靠近，这意味着所有字段在同一 Web 页面上彼此靠近

在测试问题中，我们将在本文中研究目标关系  $R$  为网络上出现的书籍作者标题对的集合。显然，这不是很好的定义，但是给定了潜在的作者和标题，以及在网络上被人提及的地方通常可以判断这是否合法书

如果我们计算  $R$  的近似值，则覆盖范围为  $\frac{|R|}{|D|}$  和错误率是  $\frac{|D - R|}{|D|}$ 。我们的目标是最大化覆盖范围并最大程度地减少错误。但是，低错误率比高覆盖率更重要。

足够大的数据库  $D$ ，只要召回就可以了，但是错误率对于许多应用程序可能是无用的

通常，我们实际上无法计算  $R$ ，因此我们无法知道 coverage 和差错率的精确值。但是，我们可以通过让用户检查  $R$  的随机元素来采样差错率。Coverage 很难估计

### 模式

直观地讲，模式匹配目标关系元组出现的一种特定格式。理想情况下，该模式要足够具体，以至于不匹配该关系中不存在的任何元组，但是在实践中，可能会出现一些误报。模式可能具有多种表示形式我们使用了非常有限的一类正则表达式

令  $p$  为模式，则  $MD(p)$  为与  $D$  中的  $p$  匹配的元组集合， $|MD(p)|$  为  $MD(p)$  中的元素数。然后， $p$  的覆盖范围和  $p$  的错误率为  $\frac{|MD(p)|}{|D|}$  和  $\frac{|D - MD(p)|}{|D|}$

对于一组模式  $P = \{p_1, \dots, p_k\}$ , 我们定义为  $MD(P)$   
 类似地扩展  $CD(P, R)$  和  $ED(P, R)$   $MD(P)$  的替代定义  
 可能需要一个元组来匹配多种模式, 请参见

$S_{p \in MD(P)}$  我们

### 模式关系对偶

一个重要的观察结果是, 给定一组具有高覆盖率和低错误率的模式  $P$ ,  
 我们可以简单地通过找到与所有模式的所有匹配项来构造一个非常好的  
 $R$  近似值。元组但是, 我们也希望给定一组良好的元组以相反的属性,  
 我们可以构建一组良好的模式。我们可以通过找到  $D$  中所有元组的出现  
 并发现出现的相似性来做到这一点。来自模式的元组和来自元组的模式  
 赋予我们强大的功能, 并且是我们在本文中提出的技术的基础

### 双重迭代模式关系提取

双重迭代模式关系提取 DIPRE 是一种利用模式关系对偶性提取关系的技  
 术, 其工作原理如下:

#### $R$ 样本

从目标关系的一个小样本  $R$  开始。该样本由用户提供, 可能很小。在  
 我们的测试中, 我们使用了五本书的作者列表

#### 发现事件 $RD$

然后发现  $D$  中所有  $R$  元组的出现。在我们的实验中, 这些是作者和书  
 名附近的出现以及文本中书名的发现, 以及找到的元组保留了每个出  
 现的 URL 和周围文本的上下文。

#### $P$ 基因型 0

根据事件集生成模式这是算法的棘手部分粗略地说, 此例程必须为具有  
 相似上下文的事件集生成模式。模式必须具有较低的错误率, 因此重要  
 的是它们不要过于笼统。模式的覆盖范围越广, 效果越好。但是, 较大  
 的数据库可以弥补覆盖范围较小的问题

$R \cap MD(P)$  在数据库中搜索与任何模式匹配的元组

如果  $R$  足够大, 则返回 Else 进入步骤

### 控制扩展

上面的过程不一定很稳定, 可能会偏离  $R$ 。特别是  $MD(P)$  中的几个伪元组会  
 在下一次迭代中导致  $P$  中的几个伪模式, 这又会导致整个伪元组变斜。

因此，GenPatterns 例程必须谨慎，以将潜在的伪元组或几个小元组所造成的损害最小化。另一安全措施是更严格地定义 MD P，以便要求元组匹配 P 中的多个模式。在我们执行的测试中不是必需的，但在以后的测试中可能是必需的。最后，随着关系的扩展，可能需要调整各种阈值

## 寻找作者和标题

对于我们的实验，我们选择从万维网上计算作者标题对的关系。这个问题非常适合 DIPRE，因为许多网站上列出了许多知名书籍，许多网站都符合整个站点的格式合理

### 1 种书籍图案

为了使用 DIPRE 来查找书籍，有必要定义什么模式包括：模式的定义在很大程度上决定了 DIPRE 的成功。但是，对于我们的测试，我们使用了非常简单的模式定义，这需要进一步研究以确定更复杂的图案定义是否效果更好

我们将模式定义为一个元组  $urlprefix\ prefix\ middle\ suffix$ ，其中  $order$  是一个布尔值，其他属性是字符串如果  $order$  为  $true$ ，则如果集合中有一个文档，则作者标题对将与该模式匹配  $WWW$  具有与  $urlprefix$  匹配且包含与正则表达式匹配的文本的 URL

前缀作者中间标题后缀

作者仅限于

$[AZ\ [A\ Za\ z\ \dots i]^{2,30}[A\ Za\ z]$

标题仅限于

$[AZ\ [A\ Za\ z\ \dots i]^{2,30}[A\ Za\ z]$

如果命令为假，则标题和作者切换为 1 发生次

数

我们还必须定义事件的结构，因为它应该有一个与模式定义的对关系作者标题对的出现由七个元组组成

作者标题顺序  $url$  前置  $x$  中间  $suffix$

该顺序与标题和作者在文本中出现的顺序相对应。 $url$  是其出现在文档上的 URL。 $prefix$  由测试中的  $m$  个字符组成  $m$  在作者或标题之前，如果标题为

*rst* 中间是作者和标题之间的文本，而 *su x* 由标题或作者之后的 *m* 个字符组成

## 1 书籍生成模式

DIPRE 程序的重要组成部分是 *GenPatterns* 例程，该例程将出现的一组书籍转换为模式列表。这是一个不平凡的问题，并且模式识别的全部领域都致力于解决该问题的一般版本出于我们的目的，我们使用一组简单的启发式方法从事件中生成模式，只要很少有误报模式会生成非图书，就足够了。每个模式只需要覆盖很小的范围，因为网络非常庞大，并且来源很多信息，因此所有模式的总覆盖范围仍然很大

假设我们给出了一组事件，并且我们希望构建一个与所有事件都匹配的特定模式，我们可以按照以下步骤进行操作

验证所有出现的顺序和中间是否相同如果不是，则不可能生成将它们全部匹配的模式分别将输出模式顺序和中间模式设置为顺序和中间  
查找所有 *url* 中最长的匹配 *pre x* 将 *outpattern urlpre x* 设置为该 *pre x*  
将 *outpattern pre x* 设置为出现的 *pre xs* 的最长匹配 *su x*  
将 *outpattern su x* 设置为出现的 *su xs* 的最长匹配 *pre x*

我们将这个例程表示为 *GenOnePattern 0*

模式特定的城市上面生成的模式可能太笼统或太具体，我们不担心它过于具体，因为会生成很多模式并合并在一起会有很多书籍，但是该模式可能太笼统，可能产生许多非书籍

为了解决这个问题，我们尝试测量模式的特定城市。模式 *p* 的特定城市大致对应于  $\log PX MD p$ ，其中 *X* 是在 *R* 的元组域上均匀分布的一些随机变量。为了快速计算，我们使用了以下内容模式 *jsj* 的特定城市的公式表示 *s* 的长度

特殊城市 *p jp middlejjp urlpre xjjp pre xjjp su xj*

---

· 如果行的结尾或开始处接近行，则 *pre x* 和 *su x* 实际上可能少于 *m* 个字符，但这是当前实现的限制，目前尚不清楚它是正数还是负数  
如果该域像所有字符串的空间一样在 *nite* 中，则均匀分布可能不明智，应使用不同的分布

我们拒绝指定城市过低的任何模式，以免产生过于笼统的模式。更具体地讲，我们坚持要求指定城市  $pnt$ ，其中  $n$  是出现支持模式  $p$  和  $t$  的书数，这是一个阈值。模式的字符串是非空的，否则指定城市为零。我们还要求  $n$ ，因为将模式基于一个示例非常容易出错

生成模式的算法这里我们介绍 Gen 模式的算法 0，它利用了在本节中介绍的 GenOnePattern 0 算法

将 0 中所有出现的  $o$  按顺序和中间分组，使结果组为  $0\ 0k$   
对于每个组  $0i\ p\ \text{GenOnePattern}\ 0i$  如果  $p$  符合特定城市  
    然后输出  $p$  否则  
    如果  $0i$  中的所有  $o$  具有相同的 URL，则拒绝  $0i$   
    否则，将  $0i$  中出现的  $o$  分成按 URL 中的字符分组的子组，这是过去的  
    一个  $urlpre\ x$  重复此过程  
    在这些亚组中

当生成的模式不够明确时，此例程使用基于 URL 的一个简单的进一步细分，也可以想象使用  $pre\ x$  或  $su\ x$

我们已经描述了一种从事件簿列表中生成模式的简单技术，人们可以想象到更复杂的技术，这是进一步研究的主题。但是，正如结果所表明的，即使这种简单的方案也能很好地工作

## 1.1 性能问题

DIPRE 有两个非常艰巨的任务：给定一长串的书，查找书本的出现；给定一个模式的清单，查找模式的匹配这两个操作都必须在非常大的 Web 文档数据库上进行

对于查找书本的第一个任务，我们首先通过两个 *fgrep* 过滤器传递数据，一个仅通过包含有效作者的行，另一个仅通过包含有效标题的行，之后是编写程序的任务 Python 实际检查该行中是否有匹配的作者和标题，以标识它们并产生出现的结果作为输出，为此，尝试了几种在 Flex 和 Python 中涉及大型正则表达式的替代方法，但它们很快超出了各种内部界限

对于第二个任务，我们只使用一个 Python 程序。每个模式都转换为一对正则表达式，一个用于 URL，一个用于实际出现。每个 URL 首先经过测试，以查看适用于其的模式，然后程序对每一行进行测试。相关的正则表达式这种方法相当慢，需要改进。将来的版本可能会使用 Flex 或

rex C 库，通过仅定位与模式匹配的 URL 可以使此任务稍微容易一些，我们为此做了一些尝试。但是，如何构造数据却使其变得不那么琐碎，我们希望我们开发出的技术能够足够通用，无法处理对 URL 的任何限制

在此时，从事件中生成模式并不是很大的性能问题，因为仅生成了数千个事件。运行较大的测试时，这将变得更加重要。当前，使用 gsort 按顺序和中间对事件进行排序，然后 Python 程序读取结果列表并生成模式，如本节所述

## 实验

尽管由于时间限制，到目前为止进行的实验非常有限，但它们却产生了非常积极的结果。

### 实验中使用的网络数据

对于数据，我们使用了总计 GB 的数百万个 Web 页面的存储库。此数据是 Stanford WebBase 的一部分，用于 Google 搜索引擎 BP 和其他研究项目。作为搜索引擎的一部分，我们建立了整个存储库的反向索引

该存储库跨越许多磁盘和几台机器，即使不进行任何实质性处理，也仅花费相当多的时间对数据进行一次传递，因此在这些情况下，我们仅在任何给定的迭代中对存储库的子集进行传递

此项目的重要说明是，存储库几乎不包含来自 Amazon Ama 的网页，这是因为它们的自动生成的 URL 使得难以抓取

### 模式关系扩展

艾萨克 • 阿西莫夫 (Isaac Asimov)	黎明机器人
大卫 • 布林	惊涛骇浪
詹姆斯 • 格里克	混沌创造了新科学 Charles
Dickens	寄予厚望
威廉 • 莎士比亚	永恒喜剧

无花果书籍的初始样本

URL 格式	模式图式
www sff 净轨迹 c *	<LI> <B>标题<B>作者
dns city net com lmann 授予 hugos html <i> title <i>作者	
海豚 upenn 教育 dcummins 义本 st 奖 htm 作者标题	

图 3 第一次迭代中发现的模式

我们仅用书籍开始实验，请参见图  
 发生和生成的模式，请参见图。  
 前五本书中的两本产生了模式，因为它们都是科学书籍。在匹配的 URL 上运行这些模式产生了独特的作者标题对。它们大多是科学的，但也有些例外，请参见图。

高清埃弗里特	死亡面具和其他鬼魂
HG 威尔斯	月球第一人
HG 威尔斯	科幻小说卷 3
HG 威尔斯	HG Wells 月球上的第一批
人	看不见的人
HG 威尔斯	莫罗博士岛
HG 威尔斯	科幻小说卷 J
HG 威尔斯	事物的形状即将来临终极革命 HG 威尔斯
	时间机器
HG 威尔斯	世界大战
HG 威尔斯	当卧铺醒来时
胡佛爵士	空荡荡的旅程 HP
Lovecraft August Derleth	门槛的潜伏者
惠普 Lovecraft	在疯狂山脉和其他恐怖故事中 HP Lovecraft
	查尔斯•德克斯特•沃德案
Lovecraft	疯狂山脉和其他恐怖故事

图第一次迭代中发现的书籍样本

在大约一百万个网页中进行搜索后发现这些书的出现是令人失望的，因为它不是第一次迭代中的大爆炸，但是整个库至少要花几天的时间因此我们没有尝试产生更多这些出现的形式，其 URL 前缀不是完整的 URL 超过两百万个 URL 产生了独特的作者标题对。不幸的是，其中有些伪造的书尤其是合法的标题但是有结论的作者。我们从列表中删除了这些。这是整个过程中唯一的人工干预。在以后的实验中，



看看是否保留这些东西会产生大量垃圾是很有趣的

对于最终迭代，我们选择使用包含工作簿的存储库的子集。该子集由大致的文档组成扫描产生的剩余书籍这些生成的模式依次扫描同一组文档产生的唯一书籍很少伪数据参见图

这项实验正在进行中，希望很快会产生更多的书籍清单。

### 结果质量

为了分析结果的质量，我们从列表中随机选择了 20 本书，并尝试通过在 Amazon Ama Visa 购物指南中搜索书籍（即斯坦福在线图书馆目录和网络）来验证它们是否是真实书籍。结果的质量都是真实的书，其余的书实际上是一篇文章

为什么我为 Andrew Carias 投票给用户用车

令人惊讶的是，除了网络之外，在某些或所有来源中都找不到大量的书籍，其中一些是在线书籍：有些晦涩或绝版：一些根本没有在某些网站上列出，没有明显的原因总共不是在亚马逊上的书籍，后者声称拥有数百万本书的目录

除了上面提到的文章外，数据还存在一些明显的问题。由于差异很小，因此多次提到了一些书，例如大写间距，作者的列出方式例如 ER Burroughs 与 Edgar Rice Burroughs，但是幸运的是，作者对于如何他们的名字被列出，并且这些重复是有限的。在某些情况下，一些信息被附加到作者的名字上，例如出版日期。

### 结论

我们的总体目标是能够利用其庞大的规模从整个万维网中提取结构化数据。已证明 DIPRE 是查找书单的简单示例中的出色工具。它从一整套书本开始并进行了扩展只需很少的人工干预，就可以得到较高质量的书本清单。同一工具可以应用于许多其他领域，例如电影，音乐餐厅等。该工具的更高级版本可能能够提取人员目录产品目录及更多

---

不幸的是，在进行这些测试时，国会图书馆搜索系统已关闭

亨利 ● 詹姆斯	欧洲人
亨利 ● 詹姆斯	金碗
亨利 ● 詹姆斯	一位女士的画像
亨利 ● 詹姆斯	螺丝的转动
亨利 ● 詹姆斯	螺丝转动
亨利 ● 约翰 ● 可口可乐	滚石的痕迹
亨利 ● K ● 罗	基督教历史上的地标
亨利 ● 基索尔	和风
亨利 ● 劳森	在世界广阔的日子里
亨利 ● 朗费罗	Hiawatha 之歌
亨利 ● 米勒	热带癌症
亨利 ● 彼得罗斯基	设计发明
亨利 ● 彼得罗斯基	有用事物的演变
亨利 ● 罗斯	叫它睡觉
亨利 ● 萨姆纳 ● 缅因州	古代法律
亨利 ● 塔克曼 ● 林赛 ● 菲拉 (Henry Tuckerman Lindsay Phila)	文学的特征亨利 ● 范 ● 戴克 蓝花
亨利 ● 范 ● 戴克 ● 斯克里布	天 0
亨利 ● 范 ● 隆	彼得 ● 史蒂文森的生活和时代
Henry Wadsworth Longfellow	保罗 ● 里维尔的骑行
亨利 ● 沃兹沃思 ● 朗费罗	伊万杰琳
亨利 ● 沃兹沃思 ● 朗费罗	Hiawatha 赫伯
特 ● 唐纳德之歌	林肯
赫伯特 ● 哈特	西北古堡
小赫伯特 ● 梅森	拉斐特 Escadrille
赫伯特 ● 洛特曼	朱尔斯 ● 凡尔纳探索传记
赫伯特 ● 斯宾塞	人与国家
赫尔曼 ● 戴利	为了共同利益
赫尔曼 ● 戴利	珍视地球
赫尔曼 ● 基特雷奇	英格索尔 (Ingersoll) 传记欣赏
赫尔曼 ● 哈肯 (Herman Haken)	脑功能原理
赫尔曼 ● 黑塞	德米安
赫尔曼 ● 黑塞	悉达多
赫尔曼 ● 黑塞	西达塔
赫尔曼 ● 梅尔维尔	巴特比
赫尔曼 ● 梅尔维尔	比利 ● 巴德
赫尔曼 ● 梅尔维尔	比利 ● 巴德
赫尔曼 ● 梅尔维尔	白鲸迪克
赫尔曼 ● 梅尔维尔	自信的人
赫尔曼 ● 梅尔维尔	妖精或魔法岛
赫尔曼 ● 梅尔维尔	偷窥波利尼西亚生活
赫尔曼 ● 魏斯	日落侦探
赫尔曼 ● 伍克 (Herman Wouk)	战争与纪念
赫尔曼 ● 黑塞	克林索尔的去年夏天
赫尔曼 ● 黑塞	努尔
赫尔曼 ● 黑塞	罗斯哈尔德
赫尔曼 ● 黑塞	来自另一颗星的奇怪新闻
希罗多德斯	历史沿革
希罗多德斯	历史
希罗多德斯	希罗多德的历史
赫歇尔 ● 霍布斯	牧师手册
黑舍尔	第一阶段的月亮
希尔森	暴风雨的天气
希莱尔	节日与新来港定居人士
希莱尔	大异端
希拉里 ● 贝利 (Hilary Bailey)	卡桑德拉 ● 特洛伊公主
希拉里 ● 诺曼 (Hilary Norman)	苏珊娜的钥匙
希尔伯特 ● 申克	计时序列
希尔伯特 ● 申克	阿巴科礁石之战
希尔达 ● 康克林	小女孩的诗
希尔达 ● 休斯	颤科
希尔达 ● 休斯	当墓地打哈欠
希勒曼	时间小偷
希勒曼	皮肤行者
希勒曼	说话的上帝
希拉姆 ● 科森 (Hiram Corson)	勃朗宁介绍
哈贾马尔霍乔斯博伊森	挪威的少年时代
哈贾马尔哈乔斯博伊森	来自两个半球的故事

无花果纳尔列表中的书样本

## 可伸缩性和稳定状态

此方法的可伸缩性面临几个挑战，一个是在巨大的存储库中扫描大量模式和元组所需的性能，底层算法和实现的改进很可能在不久的将来解决此问题。

一个潜在的更大难题是，在扩大关系时是否可以防止 DIPRE 与目标脱节？例如，由于 DIPRE 实际上只使用了种子样本中的两本科学书籍，为什么它没有产生大量的科学显然，它倾向于所有书籍的汇编，甚至是一些零散的文章都没法建立了这种关系，随着关系的扩展而保持这种效果是不平凡的，但是有几种可能性

与奇异值分解的连接一种可能性是重做

MD P 的  $ne$  要求多个模式匹配一个元组一个更极端的版本是为每个元组和每个模式分配一个权重。一个匹配的元组根据它匹配的模式权重分配一个权重。基于匹配的元组的权重的权重如果以线性方式完成此技术，则分解为元组模式矩阵的奇异值分解乘以其转置。这类似于对文档单词执行的潜在语义索引 DDF 矩阵在这种情况下，最终的稳态是主导特征向量。幸运的是，它与初始样本无关，这显然是不可取的。尽管如此，它与 LSI 的关系是令人信服的，需要进一步研究。

即使不使用权重，稳态与上述初始状态的独立性也可能是一个问题。有几种可能的解决方案之一，就是像我们在本文中演示的那样，仅通过有限数量的迭代来运行。另一种解决方案是确保元组到模式到元组的转换是非线性的，并且具有一些依赖于初始状态的局部稳态。这可以通过在 GenPatterns 的计算中使用初始样本  $R$  来完成。在这种情况下，用户还可以提供  $R$  反例列表

## 自动提取的含义

该实验最令人惊讶的结果之一是查找未在主要在线资源中列出的书，例如在线出版的 Douglas Clark Cla 的《Disbanded》或 Dollie Radford Rad 出版的《The Young Gardeners Kalendar》，这本晦涩难懂的书在 If If book 中出版。可以扩展列表，并且如果可以提取在线资源中列出的几乎所有图书，则结果列表可能比任何现有图书数据库都更完整。生成的列表将是成千上万个小型在线资源的产物，例如

与当前书籍数据库相对，后者是一些大量信息源的产物。这种信息流的变化可能会产生重要的社会影响

## 参考文献

- 阿玛 亚马逊首页 <http://www.amazon.com>
- 英国石油公司 谢尔盖·布林和拉里·佩奇 谷歌搜索引擎 谷歌斯坦福大学
- 布里 谢尔盖·布林 书籍清单 <http://www.db.stanford.edu/谢尔盖书本.html>
- 克拉 道格拉斯·克拉克 (Douglas Clark) 解散了本杰明·媒体 (Benjamin Press) Hillcrest Drive 浴 Ba3 JHD 英国 <http://www.bath.ac.uk/exxdgdc/诗歌库/di.html>
- 敌敌+ 潜在语义分析对 Scott Deerwester 的 Susan Dumais Goerge 畏 Furnas Thomas Landauer 和 Richard Harshman 进行了索引/美国信息科学学会杂志 JJJ
- 关于半结构化数据管理的 MOS 研讨会 <http://www.attcom.com/suciu/研讨会论文.html> 五月 J
- 拉德 多莉·雷德福德 (Dollie Radford) 年轻的园丁卡伦达尔 亚历山大·莫尔 ing Ltd 伦敦 J <http://www.印第安纳州教育人士.vwwp.radford.kalendar.html>
- si 尖沙咀主页 <http://www.db.stanford.edu/尖沙咀尖沙咀.HTML> 格式
- 可见 Visa 书籍购物指南 <http://商店指南.yahoo.com/商店指南书籍.html>