

基于旅游领域本体和 Nutch 的泛北部湾旅游资源搜索引擎的应用研究

◆陈意山

(广西师范大学漓江学院 广西桂林 541006)

摘要: 本体在自然语言处理、人工智能的最新领域使用广泛。本文尝试对旅游领域本体的属性和关系的描述,并通过基于 Nutch 建立的搜索引擎,在面向北部湾旅游资源,可根据用户的要求进行排列输出和本地化查询,提高了对泛北部湾旅游资源信息检索的效率。

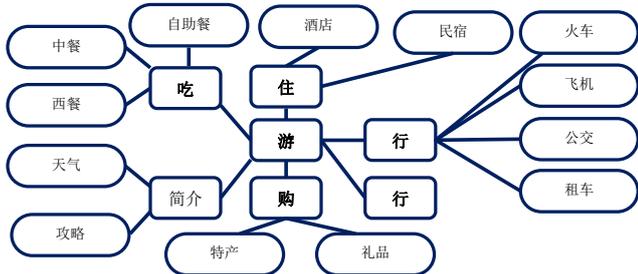
关键词: 领域本体; 旅游; 搜索引擎; Nutch; 泛北部湾

浏览网页是旅行者获取旅游信息的最主要来源,旅行者经常需要手工查找筛选,而通过传统搜索引擎进行信息检索会找到成千上万的网页,而真正有帮助的信息需要在这些大量纷繁复杂的网页仔细寻找,因此需要针对特定范围的旅游搜索引擎的开发势在必行。

1 相关技术

1.1 旅游领域本体构建

本体本质上是概念及其关系之间的模型,必须强调的是这些概念和概念之间的关系必须是能够被描述、形式化的。这样较为深刻的内涵知识就能够通过本体描述的概念及其关系体现出来。本系统将首先阐述基于旅游领域知识本体的领域概念,旅游领域本体概念包括吃、住、行、游、购、娱、简介这几项,在建立旅游领域本体数据库时,基于测试和数据量的角度,数据主要来源于泛北部湾地区的旅游资源,其本体库如下图。



作为旅游的另一个主体——游客,也需要构建游客本体。游客在旅游过程中需要和旅游的六大主要元素进行交互,为了提高资源搜索的效率和智能化,游客本体常见属性为性别、年龄段、工作、学历、兴趣、婚姻状况等信息。

1.2 Nutch

Nutch 是基于 Java 语言的开源搜索引擎,可为开发人员自行配置符合自己需求的搜索引擎和相关工具。Nutch 包括爬虫 Crawler 和查询 Searcher 两个模块构成。Crawler 的功能是从互联网上抓取网页并建立相应的网页索引。Searcher 的功能是利用 Crawler 建立的索引检索游客输入的的查找关键词来产生查找结果。这两个模块相互关联只有索引文件,因此两模块耦合程度较低。

1.3 Jena 推理机

Jena 是美国惠普实验室设计的开放式 Java 语言框架工具包,它是一种面向本体的、比较成熟的推理机,拥有对本体进行解析、存储、推理和查询的函数调用和处理接口。因此本文选择使用该推理机结合为 Java 平台的开源搜索引擎 Nutch 实现本系统所要完成的需求。

2 系统需求分析

2.1 系统基本功能: 用户登录到发布到 Tomcat 的网页上,输入想搜索的旅游资源关键字后,搜索引擎会返回给用户一系列包含用户输入的关键字的网页地址,网页标题,以及网页摘要等,用户可以从显示的查询结果信息中选中一个标题,打开就可以浏览所选定的网页信息。

2.2 系统模块组成

(1)网络爬虫:任务是从目标源定时抓取信息资源。

(2)用户接口:用户接口的作用是输入用户查询,显示查询结果,提示用户相关性反馈机制。用户接口的主要目的是方便用户使用搜索引擎,高效率,多方式地从搜索引擎中得到有效,及时的信息。

(3)索引:索引的作用是理解搜索引擎所搜索的信息含义,然后对爬取的网页进行内容分析,从中抽取索引项并获得相关的网页信息。

2.3 索引模块设计实现

仅仅建设了旅游领域本体和游客领域本体还是不够的,需要对旅游信息本体数据库进行更为精确的搜索查询才能满足系统的需求,因此本系统利用 Jena 推理机来提供搜索精确度和效率。

首先需要使用语义精确方法与旅游领域本体中的概念相关联,建设满足一致性的公理库。根据用户提交的查询请求,根据该请求生成面向旅游本体的查询语句。然后基于该公理库的数据,利用 Jena 推理机对知识库中的知识进行检查和推理,再生成推理查询结果反馈给用户。从而实现了旅游领域知识之间的信息联通,能够为游客提供更为精确和高效的旅游信息。

3 结束语

更为智能的旅游搜索工具将是未来旅游行业技术发展的一个方向,本文介绍的基于旅游领域本体和 Nutch 的旅游信息搜索引擎的设计在功能上还有一定缺陷,要改进的地方确实很多。但是经过测试,对比的搜索引擎均为全网搜索,搜索结果中会出现大量与泛北部湾旅游资源无关的内容,这就是本搜索引擎有优势的地方。

参考文献:

- [1]王静,刘伟峰,汪伟.面向旅游信息的垂直搜索引擎的设计与实现[J].北京:信息系统工程,2014(3):29-31.
- [2]冯欣,王成良.本体在旅游信息系统中的应用研究[J].北京:计算机与现代化,2010(3):128-132.
- [3]吴起立.基于旅游领域本体的自动分类构建研究[J]经济研究导刊,2013(27):274-275.
- [4]谭月辉,肖冰等.Jena 推理机制及应用研究[J]河北省科学院学报,2009(26):14-17.

作者简介: 陈意山(1977年5月-),男,广西玉林人,硕士,副教授,从事智能教学系统研究。

基金项目: 广西师范大学漓江学院院级科研项目(基于 QTI 标准的人体动作描述研究)。

