

<<语义桌面搜索技术>>

图书基本信息

书名：<<语义桌面搜索技术>>

13位ISBN编号：9787562935964

10位ISBN编号：7562935963

出版时间：2011-9

出版时间：武汉理工大学出版社

作者：李胜

页数：166

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<语义桌面搜索技术>>

内容概要

要对桌面文档进行搜索，需要某种信息检索模型来支持，而信息检索模型一直都是信息检索领域的重要研究课题。

在对传统向量空间模型研究的基础上，《语义桌面搜索技术》设计了一种基于本体的语义信息检索模型，其工作重点包括：语义项权重的设计，各关键字之间的语义关系分析，以及语义特征向量之间的相似度计算策略等。

在模型中，通过概念连通刚对不同语义项之间的关系进行了重新考量，并将语义相似度的计算分为概念相似度和属性相似度两个方面，综合考虑了二者在语义检索中的作用，改善了检索效果。

对检索结果进行排序是文档检索的一个重要步骤。

书中在对现有的web排序算法和模式图理论进行研究后，提出一种基于权威传递的检索结果排序方法

。该方法利用本体描述文档之间权威传递的模式，通过设定不同的传递系数，反映出文档之间的不同连接线索对文档联系紧密程度的影响。

实验结果表明，该方法可将重要程度高的结果优先返回，并有效地体现出文档对象之间的关联关系。

<<语义桌面搜索技术>>

书籍目录

- 1 绪论
 - 1.1 语义桌面研究的背景和意义
 - 1.2 桌面搜索的需求
 - 1.3 本文研究的内容
 - 1.4 文章的组织结构
- 2 语义网技术
 - 2.1 语义网
 - 2.2 基于语义网的检索技术
 - 2.3 语义搜索模型
 - 2.4 研究方法总结
- 3 语义桌面
 - 3.1 语义桌面的产生和发展
 - 3.2 语义桌面的体系结构及其组成部分
 - 3.3 语义桌面的研究现状
 - 3.4 语义桌面搜索
 - 3.5 当前语义桌面搜索中面临的问题
 - 3.6 小结
- 4 基于用户行为的桌面元数据提取
 - 4.1 桌面元数据
 - 4.2 基于用户行为的元数据提取
 - 4.3 桌面文档本体的创建
 - 4.4 动态上下文检测方法
 - 4.5 元数据生成器原型系统实现
 - 4.6 小结
- 5 桌面无结构文档的实体提取
 - 5.1 信息提取概述
 - 5.2 无结构文档的实体提取
 - 5.3 基于本体的无结构文档实体识别方法
 - 5.4 基于PLSA的无结构文档关系识别方法
 - 5.5 小结
- 6 桌面文档检索模型
 - 6.1 向量空间模型概述
 - 6.2 语义向量空间模型的新特点
 - 6.3 创建本体
 - 6.4 语义向量的相似度计算
 - 6.5 实验分析
 - 6.6 小结
- 7 文档的语义排序方法
 - 7.1 语义排序
 - 7.2 基于权威传递的排序方法
 - 7.3 算法的时间效率分析
 - 7.4 实验及效率评估
 - 7.5 小结
- 8 总结与展望
 - 8.1 总结

<<语义桌面搜索技术>>

8.2 今后的展望
参考文献

<<语义桌面搜索技术>>

章节摘录

版权页：插图：Web浏览器缓存中蕴涵了大量的用户浏览行为信息，这些信息不但对查找相关的结果很有帮助，而且能够为搜索结果提供更多的扩展上下文信息。

例如，我们在桌面上搜索到某篇论文的文档，该文档是通过网页下载得到的。

通过Web缓存上下文元数据，就可以得到这篇论文的参考文献信息，以及与该论文议题相关的其他论文的URL信息。

通过研究我们发现，人们习惯于将事物关联到某个特定的上下文中，所有这些上下文信息在搜索时都有可能被利用到。

然而到目前为止，桌面搜索工具既没有将这些信息与文档关联起来，也没有在搜索中使用到它们。

4.1.2桌面元数据的作用 在现有的语义桌面研究项目中，元数据起着非常关键的作用。

综合来看，至少可以分为以下三个方面：1.利用元数据扩展搜索结果 这个方面的代表就是TAP，TAP是一项著名的语义搜索工程。

建立TAP的目的有两点：一是把来自不同Web服务上的数据整合成一个统一的全局数据库，于是就建立了TAP知识库；二是为知识库提供接口便于数据的访问，如GetData等。

TAP Knowledge Base是一个浅显的但内容广泛的知识库，涉及的领域很多，包括人物、组织、地理和产品等。

它是R.Guha等人开发的语义搜索系统ABS的一个重要的知识来源。

该知识库以RDF作为描述资源及其内部关系的数据模型。

GetData是一个简单的查询接口。

每个GetData查询都是声明与该URL相连的SOAP消息。

此消息包括两个参数：被访问属性的资源和被访问的属性。

GetData查询返回的结果是语义网上知识库内部包含资源的图，还有所查询资源和各来源之间的关系。

用抽象语法表示如下：GetData(,) (value) TAP建立在TAPache模块之上，它提供一个平台来发布和使用语义网上的数据。

TAP的一个辅助系统onTAP对知识库提供更新服务，包括207个HTML页面模板，可以从38个高质量网站中读取和提取数据。

TAP的核心思想是：对于特殊搜索而言，一些信息对于编制目录和后端数据库十分有用，但是他们却无法被Google网络爬虫那样的工具访问到。

基于语义搜索的结果将扩展通过传统技术得到的检索结果，同时这些语义搜索结果又是独立存在的。

主题分类是网页非常重要的特征，当用户要搜索某个著名实体的时候（例如城市、国家），可以利用Web中的许多信息，其中就包括页面的主题分类。

这些信息在后端数据库中是十分有用的，而这些数据集合可以与实体的ID联系起来。

<<语义桌面搜索技术>>

编辑推荐

《语义桌面搜索技术》在传统信息提取技术的基础上，给出了一种基于本体的无结构文档信息提取方案。

书中在对现有的Web排序算法和模式图理论进行研究后，提出一种基于权威传递的检索结果排序方法。

。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>