

[本期目录](#) | [下期目录](#) | [过刊浏览](#) | [高级检索](#)

[\[打印本页\]](#) [\[关闭\]](#)

## 软件技术与数据库

### 基于本体的Deep Web数据源发现方法

李道申, 刘勇

(河南科技大学电子信息工程学院, 河南 洛阳 471003)

**摘要:** 提出一种基于本体的Deep Web数据源发现方法, 采用网页分类、表单内容分类、表单结构分类方式, 确定符合某领域的Deep Web查询接口。在网页分类和表单内容分类中引入本体的半自动构建和自动扩展模块, 在表单结构分类中添加启发式规则。实验结果证明, 该方法能有效提高Deep Web数据源的查全率和查准率。

**关键词:** 深网 本体 数据源 半自动构建 分类模型

### Deep Web Data Sources Discovery Method Based on Ontology

LI Dao-shen, LIU Yong

(College of Electronic Information Engineering, Henan University of Science and Technology, Luoyang 471003, China)

**Abstract:** This paper presents a Deep Web data sources discovery method based on ontology. It uses webpage classification, form structure classification and form content classification to find Deep Web querying interface in some fields. It proposes that semi-automatic construction and automatic extension of ontology are added to the webpage and form content classification, and heuristic rules are enriched in the form structure classification. Experimental results show that this method can improve the precision and recall of Deep Web database discovery effectively.

**Keywords:** Deep Web ontology data sources semi-automatic construction classification model

收稿日期 2011-07-19 修回日期 网络版发布日期 2012-02-20

DOI: 10.3969/j.issn.1000-3428.2012.04.017

基金项目:


国家自然科学基金资助项目(70671035)

通讯作者:

作者简介: 李道申(1986—), 男, 硕士研究生, 主研方向: Web数据挖掘; 刘勇, 教授

通讯作者E-mail: lidaoshen.good@163.com

## 参考文献:

- [1] 马建华, 李赛红, 徐兰兰. 深层网中基于入口查询的表单填充策略[J]. 计算机工程. 2010, 36(7): 66-67 [浏览](#)
- [4] 王辉, 刘艳威, 左万利. 使用分类器自动发现特定领域的深度网入口[J]. 软件学报. 2008, 19(2): 246-256 
- [5] 周子力, 顾君忠. WordNet的本体构建及其在安全领域应用关键技术研究[D]. 济南: 山东师范大学, 2009.

## 扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(290KB\)](#)
- ▶ [\[HTML\] 下载](#)
- ▶ [参考文献\[PDF\]](#)
- ▶ [参考文献](#)

## 服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [引用本文](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

## 本文关键词相关文章

- ▶ [深网](#)
- ▶ [本体](#)
- ▶ [数据源](#)
- ▶ [半自动构建](#)
- ▶ [分类模型](#)

## 本文作者相关文章

- ▶ [李道申](#)
- ▶ [刘勇](#)

## PubMed

- ▶ [Article by Li, D. S.](#)
- ▶ [Article by Liu, Y.](#)

[6] 马 军, 宋 玲, 韩晓晖, 等. 基于网页上下文的Deep Web数据库分类[J]. 软件学报. 2008, 19(2): 267-274. 

[7] 王 英, 左万利. Deep Web数据集成技术研究[D]. 长春: 吉林大学, 2010.

### 本刊中的类似文章

1. 郭建兵, 崔志明, 陈明, 赵朋朋. 基于DOM树与领域本体的Web抽取方法[J]. 计算机工程, 2012, 38(5): 56-58
2. 刘平峰, 章佩璐, 陈冬林. 电子目录本体合并中的实例消重机制[J]. 计算机工程, 2012, 38(2): 32-35
3. 陆筱霞, 李思昆, 马千里. 基于上下文的综合战场环境本体建模[J]. 计算机工程, 2012, 38(04): 272-274
4. 王倩, 王辉. 数据交换中基于本体的语义冲突消解方案[J]. 计算机工程, 2012, 38(04): 76-78
5. 徐小良, 陈金奎, 吴优. 基于聚类优化的Web服务发现方法[J]. 计算机工程, 2011, 37(9): 68-70
6. 陈宇, 朱建锋, 吴毅坚, 赵文耘. 一种基于领域本体的新术语扩充方法[J]. 计算机工程, 2011, 37(7): 24-27, 43
7. 高雅田, 李春生, 富宇. 基于关系数据分析的知识服务模型[J]. 计算机工程, 2011, 37(5): 56-58
8. 王海龙, 胡景芝, 赵朋朋, 崔志明. 基于搜索引擎的Deep Web数据源发现[J]. 计算机工程, 2011, 37(5): 77-79, 82
9. 吕金丽, 余雪丽. 课程知识本体建模及推理[J]. 计算机工程, 2011, 37(4): 61-63
10. 齐红, 张亮亮, 李昕. 基于玉米本体的语义检索系统[J]. 计算机工程, 2011, 37(4): 34-36

### 文章评论

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text" value="5173"/>
<input type="text"/>			