中文专利中本体关系获取研究

谷俊[1], 许鑫[2]

1. 上海宝山钢铁股份有限公司 上海 201900;
2. 华东师范大学商学院信息学系 上海 200241

Gu Jun[1], Xu Xin[2]

1. Baoshan Iron and Steel Co, Ltd., Shanghai 201900, China;
2. Department of Informatics, Business School, East China Normal University, Shanghai 200241, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (570KB) HTML (1KB) Export: BibTeX or EndNote (RIS) Supporting Info

摘要 介绍从中文专利摘要文本中抽取本体非分类关系的方法。首先对摘要文本的句法格式进行分析,按照“领域句式”、“特征句式”、“组件“工艺句式”和“效果句式”等构建子句抽取规则,再利用B、I、E和O等标注符号对子句中的术语进行人工标注,形成一定规模的训练语料集合,并利用CRFs实现训练语料的学习和新语料的抽取。最后给出应用实例并进行分析,验证方法的有效性。

关键词： 规则匹配 条件随机场 本体学习 非分类关系抽取

Abstract： This paper promotes a method which collects the non-taxonomic relation from the Chinese patents' texts. Firstly, it analyzes the syntax of abstract texts, then constructs the sub-sentences extraction rules by domain sentence,character sentence, module & craft sentence and effect sentence. Secondly, artificially labels the terms of sub-sentences by label symbols such as BIEO, creates a scale of training data set. Thirdly, learns the training data and extracts the new data by CRFs. Finally, analyzes the experiment results and verifies the validity of the method.

Keywords： Rule matching, CRFs, Ontology learning, Non-taxonomic relation

收稿日期: 2013-07-19;

基金资助:本文系上海市科技发展基金软科学研究项目“基于专利文献的本体构建与应用方法研究”(项目编号:13692107000)的研究成果之一。

通讯作者 谷俊 Email: frjcygu@163.com

引用本文:

谷俊, 许鑫 .中文专利中本体关系获取研究[J] 现代图书情报技术, 2013,V29(10): 73-78

Gu Jun, Xu Xin .Study on Ontology Relation Extraction in Chinese Patent Documents[J] , 2013,V29(10): 73-78

链接本文:

http://www.infotech.ac.cn/CN/ 或 http://www.infotech.ac.cn/CN/Y2013/V29/I10/73

Service

- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- Email Alert
- RSS

作者相关文章

- 谷俊
- 许鑫

[1] 邵波. 企业竞争与反竞争情报中的专利分析研究[J]. 情报科学, 2006, 24(2):235-238.(Shao Bo. The Patent Analysis of Enterprise Competitive Intelligence and Counterintelligence[J]. *Information Science*,2006,24(2):235-238.)

[2] 100004说明书摘要[EB/OL].[2012-02-13].http://www.sipo.gov.cn/bgxz/zlsqbg/ty/100004smszy.doc. (Instruction Abstract of 100004[EB/OL].[2012-02-13].http://www.sipo.gov.cn/bgxz/zlsqbg/ty/100004smszy.doc.)

[3] Girju R,Moldovan D I. Text Mining for Causal Relations[C].In: *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference*, Florida, USA.AAAI Press,2002:360-364.

[4] Byrd R J, Ravin Y. Identifying and Extracting Relations in Text[C]. In: *Proceedings of the 4th International Conference Application of Natural Language to Information System*.1999.

[5] Maedche A,Staab S. Discovering Conceptual Relations from Text[C]. In: *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000)*. 2000:321-325.

[6] 谭力,史忠植.基于数据挖掘的本体关系学习算法[J]. 郑州大学学报:理学版,2008,40(3):40-43.(Tan Li,Shi Zhongzhi. Ontology Conceptual Relation Learning Algorithm Based on Data Mining[J].*Journal of Zhengzhou University:Natural Science Edition*,2008,40(3):40-43.)

[7] 董丽丽,胡云飞,张翔.一种领域概念非分类关系的获取方法[J]. 计算机工程与应用, 2013,49(4):157-161.(Dong Lili,Hu Yunfei,Zhang Xiang. Method for Non-taxonomical Relations from Domain Concepts[J]. *Computer Engineering and Applications*,2013,49(4):157-161.)

[8] 于娟,党延忠.本体关系学习方法研究——概念特征词法[J]. 系统工程理论与实践,2012,32(7):1582-1590.(Yu Juan, Dang Yanzhong. Learning Ontology Relations from Documents:The Concept-feature Method[J]. *Systems Engineering—Theory & Practice*,2012,32(7):1582-1590.)

[9] Li L, Zhou R, Huang D. Two-phase Biomedical Named Entity Recognition Using CRFs[J]. *Computational Biology and Chemistry*,2009,33(4):334-338.

[10] Peng L, Liu Z, Zhang L. A Recognition Approach Study on Chinese Field Term Based Mutual Information/Conditional Random Fields[J]. *Procedia Engineering*,2012,29:1952-1956.

[11] Chen L, Qi L, Wang F. Comparison of Feature-level Learning Methods for Mining Online Consumer Reviews[J]. *Expert Systems with Applications*, 2012,39(10): 9588-9601.

[12] Esuli A, Marcheggiani D, Sebastiani F. An Enhanced CRFs-based System for Information Extraction from Radiology Reports[J].*Journal of Biomedical Informatics*,2013,46(3):425-435.

[13] 国家知识产权局[EB/OL].[2012-02-14]. http://www.sipo.gov.cn. (State Intellectual Property Office of the People's Republic of China[EB/OL]. [2012-02-14]. http://www.sipo.gov.cn.)

[14] ICTCLAS特色[EB/OL]. [2011-01-10]. http://ictclas.org/ictclas_feature.html. (Features of ICTCLAS[EB/OL]. [2011-01-10]. http://ictclas.org/ictclas_feature.html.)

[15] Lafferty J D, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. In: *Proceedings of the 18th International Conference on Machine Learning (ICICML'01)*, Williamstown, MA, USA. San Francisco,CA, USA: Morgan Kaufmann Publishers Inc.,2001:282-289.

[16] CRF + +: Yet Another CRF Toolkit [EB/OL].[2012-03-11].http://crfpp.googlecode.com/svn/trunk/doc/index.html.

[17] 谷俊,王昊.基于领域中文文本的术语抽取方法研究[J]. 现代图书情报技术,2011(4):29-34.(Gu Jun,Wang Hao. Study on Term Extraction on the Basis of Chinese Domain Texts[J].*New Technology of Library and Information Service*,2011(4):29-34.)

[1] 王昊, 邹杰利, 邓三鸿.面向中文图书的自动标引模型构建及实验分析[J]. 现代图书情报技术, 2013,29(7/8): 55-62

[2] 汤亚芬.先秦古汉语典籍中的人名自动识别研究[J]. 现代图书情报技术, 2013,29(7/8): 63-68

[3] 林琛, 王兰成.基于条件随机场的网民评论对象识别研究[J]. 现代图书情报技术, 2013,(6): 63-67

[4] 高强, 游宏梁.基于层叠模型的国防领域命名实体识别研究[J]. 现代图书情报技术, 2012,(11): 47-52

[5] 刘萍, 胡月红.领域本体学习方法和技术研究综述[J]. 现代图书情报技术, 2012,28(1): 19-26

[6] 王昊, 邓三鸿, 苏新宁.基于字序列标注的中文关键词抽取研究[J]. 现代图书情报技术, 2011,27(12): 39-45

[7] 逯万辉, 马建霞.基于条件随机场模型的复杂时间信息抽取研究[J]. 现代图书情报技术, 2011,27(10): 29-33

[8] 朱丹浩 王东波 谢靖.基于条件随机场的介宾结构自动识别*[J]. 现代图书情报技术, 2010,26(7/8): 79-83

[9] 郑荣廷 李楠 吉久明 滕青青.中文化学物质名称识别研究*[J]. 现代图书情报技术, 2010,26(6): 48-52

[10] 黄炜,张李义.基于语义爬虫的商品信息主题采集研究*[J]. 现代图书情报技术, 2010,26(1): 3-8

[11] 刘坤,吕学强,王涛,施水才.基于多尺度条件随机场的文本图像二值化*[J]. 现代图书情报技术, 2009,25(4): 79-81

[12] 阴晋岭,王惠临.词性标注的方法研究*——结合条件随机场和基于转换学习的方法进行词性标注[J]. 现代图书情报技术, 2009,3(3): 46-51

[13] 宗萍,施水才,王涛,吕学强.基于条件随机场的英文地理行政实体识别*[J]. 现代图书情报技术, 2009,3(2): 51-55

[14] 王东波,苏新宁.英汉双语句子级平行语料库自动构建*[J]. 现代图书情报技术, 2009,25(12): 47-51

[15] 王昊,刘建华,苏新宁,杨建林.面向语义网的本体学习技术和系统研究*[J]. 现代图书情报技术, 2009,3(1): 64-72

Copyright 2010 by 现代图书情报技术