



科技文献全文主题识别方法实证研究

叶春蕾^{1,2,3}, 冷伏海¹

1. 中国科学院国家科学图书馆 北京 100190;
2. 中国科学院研究生院 北京 100049;
3. 北京城市学院信息学部 北京 100094

Ye Chunlei^{1,2,3}, Leng Fuhai¹

1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China;
2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China;
3. Information Department, Beijing City University, Beijing 100094, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (462KB) [HTML \(1KB\)](#) Export: BibTeX or EndNote (RIS) Supporting Info

摘要 利用基于多词短语词频分析和短语邻近分析的DT方法,结合学科领域术语识别内容,对美国国家航空航天局2011-2020年战略规划做全文主题识别。实验证明,改进的DT方法可以有效地识别全文文献主题,一定程度上能够简化情报研究人员的工作。

关键词: [共词分析](#) [主题识别](#) [DT](#) [术语识别](#)

Abstract: This paper proposes an improved DT method to identify the theme of the NASA 2011-2020 strategic plan based on multi-word phrases frequency analysis and phrases proximity analysis, adding the term identification of subject. Experiment proves that the improved DT method can identify the theme of documentation in full-text effectively and simply the work of intelligences to some extent.

Keywords: [Co-word analysis](#), [Theme identification](#), [DT](#), [Term identification](#)

收稿日期: 2011-11-03;

引用本文:

叶春蕾, 冷伏海. 科技文献全文主题识别方法实证研究[J]. 现代图书情报技术, 2012, V28(1): 53-57

Ye Chunlei, Leng Fuhai. Theme Identification Empirical Study on Technical Documentation in Full-text[J], 2012, V28(1): 53-57

链接本文:

<http://www.infotech.ac.cn/CN/> 或 <http://www.infotech.ac.cn/CN/Y2012/V28/I1/53>

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

- ▶ 叶春蕾
- ▶ 冷伏海

- [1] 沈清, 胡德文, 时春. 神经网络应用技术[M]. 长沙: 国防科技大学出版社, 1993.
- [2] 谭永红. 基于BP神经网络的自适应控制[J]. 控制理论与应用, 1994, 11 (1): 84-88.
- [3] 王鑫, 于洪亮, 张琳, 等. 采用改进遗传神经网络的多载荷振动信号故障诊断[J]. 噪声与振动控制, 2011, 8: 137-141.
- [4] 徐永春, 张森文. 改进遗传神经网络在甘蔗产量预测中的应用[J]. 华南农业大学学报, 2011, 7(31): 102-104.
- [5] 高海龙, 张国立. 改进遗传神经网络及其在负荷预测中的应用[J]. 华北电力大学学报, 2009, 9(36): 37-40.
- [6] Mo Y F. A Study on Tactics for Corporate Website Development Aiming at Search Engine Optimization[C]. In: *Proceedings of the 2nd International Workshop on Education Technology and Computer Science*. 2010: 673-675.
- [7] 黄志栋, 员巧云. 基于PageRank算法的搜索引擎优化策略[J]. 情报探索, 2011(1): 34-37.
- [8] Dixon S. Live Tracking of Musical Performances Using On-line Time Warping[C]. In: *Proceedings of the 8th International Conference on Digital Audio Effects*, Madrid, Spain. 2005: 92-97.
- [9] 马晓玲, 吴永和. 对于搜索引擎优化(SEO)的研究[J]. 情报杂志, 2005, 24(12): 119-121.
- [10] 费巍, 黄如花. 基于用户行为分析的搜索引擎优化策略[J]. 图书情报工作, 2005, 49(10): 75-77.
- [11] 张振幸, 李金厚. 一种基于义原重合度的词语相似度计算[J]. 信阳师范学院学报: 自然科学版, 2010, 23(2): 296-299.
- [12] 刘群, 李素建. 基于《知网》的词汇语义相似度的计算[C]. 见: 第三汉语词汇语义学研讨会, 台北, 中国. 2002: 59-76.

- [13] Frantzi K T, Tsujii J. The C-value/NC-value Method of Automatic Recognition for Multi-word Terms[C]. In: *Proceedings of the 2nd European Conference on Research and Advanced Technology for Digital Libraries*. 1998: 585-604.
- [14] Leung F H F, Lam H K, Ling S H, et al. Tuning of the Structure and Parameters of a Neural Network Using an Improved Genetic Algorithm [J]. *IEEE Transactions on Neural Networks*,2003,14(1):79-88. 
- [15] Cont A. A Coupled Duration-focused Architecture for Realtime Music to Score Alignment[J]. *IEEE Transaction on PAMI*, 2010,32(6): 974-987.
- [16] National Aeronautics and Space Administration. 2011 NASA Strategic Plan [EB/OL]. [2011-08-05].http://www.nasa.gov/pdf/516579main_NASA2011StrategicPlan.pdf.
- [1] 杨颖, 崔雷.应用改进的共词聚类法探索医学信息学热点主题演变[J]. 现代图书情报技术, 2011,27(1): 83-87
- [2] 陆伟, 彭玉, 陈武.基于SOM的领域热点主题探测[J]. 现代图书情报技术, 2011,27(1): 63-68
- [3] 王立学, 冷伏海, 王海霞.技术成熟度及其识别方法研究*[J]. 现代图书情报技术, 2010,26(3): 58-63
- [4] 储久良.利用开源软件构建校园网络设备监控平台*[J]. 现代图书情报技术, 2010,26(2): 85-90
- [5] 许德山, 张智雄, 王峰, 邢美凤.上下文分析与统计特征相结合的英文术语抽取研究[J]. 现代图书情报技术, 2010,26(12): 28-33
- [6] 李树青, 程国达, 王维民.基于加权XML模型的XML数据与DTD模式匹配*[J]. 现代图书情报技术, 2010,26(1): 57-65
- [7] 陈仕吉.科学研究前沿探测方法综述[J]. 现代图书情报技术, 2009,(9): 28-33
- [8] 王建冬.基于复杂网络方法的国内信息服务研究概念网络分析[J]. 现代图书情报技术, 2009,(10): 56-61
- [9] 刘建华, 张智雄, 徐健, 许雁冬.自动术语识别——对科技文献进行文本挖掘的重要技术方法*[J]. 现代图书情报技术, 2008,24(8): 12-17
- [10] 岑咏华, 韩哲, 季培培.基于隐马尔科夫模型的中文术语识别研究[J]. 现代图书情报技术, 2008,24(12): 54-58
- [11] 王昊.基于层次模式匹配的命名实体识别模型[J]. 现代图书情报技术, 2007,2(5): 62-68
- [12] 陈曦, 陈新. Perl语言辅助的信息计量学研究[J]. 现代图书情报技术, 2006,1(7): 41-46
- [13] 张晗, 崔雷, 姜洋.运用非相关文献知识发现方法挖掘科研机构潜在的合作方向[J]. 现代图书情报技术, 2006,1(4): 45-48
- [14] 张健.XML文档的约束及其应用探讨[J]. 现代图书情报技术, 2005,21(5): 23-26
- [15] 高峰.MARC数据转换为XML文档的设计与实现[J]. 现代图书情报技术, 2005,21(1): 22-25