

本期目录 | 下期目录 | 过刊浏览 | 高级检索

[打印本页] [关闭]

人工智能及识别技术

基于关键词抽取的自动文摘算法?

蒋效宇

(北京服装学院商学院, 北京 100029)

摘要: 针对生成文摘内容不完整的问题, 利用相邻词的共现频率进行未登录词识别, 提出一种通过词汇链的构建进行中文关键词抽取和文摘生成的算法, 并给出一种采用《知网》为知识库构建词汇链的方法。通过计算词义相似度构建词汇链, 结合词汇所在词汇链的强度、信息熵和出现位置等属性, 进行关键词抽取和句子重要度计算。实验结果表明, 与已有算法相比, 该算法能够提高生成摘要的召回率和准确率。

关键词: 自动文摘 向量空间模型 关键词抽取 词汇链 未登录词识别

Automatic Summarization Algorithm Based on Keyword Extraction

JIANG Xiao-yu

(Business School, Beijing Institute of Fashion Technology, Beijing 100029, China)

Abstract: In order to over the shortcoming of the incomprehensive of summarization, a new lexical chain-based keywords extraction and automatic summarization algorithm from Chinese texts based on the unknown word recognition using co-occurrence of neighbor words is proposed, and an algorithm for constructing lexical chain based on Hownet knowledge database is given in the method, lexical chain is constructed by calculating the semantic similarity between terms, keywords are extracted and the importance of each sentence is calculated according to the intensity of lexical chain, the entropy of terms and position. Experimental results show that the summarization generated by the improved algorithm gets better performance than other methods both in recall and precision.

Keywords: automatic summarization vector space model keyword extraction lexical chain unknown word recognition

收稿日期 2011-06-03 修回日期 网络版发布日期 2012-02-05

DOI: 10.3969/j.issn.1000-3428.2012.03.062

基金项目:



北京市优秀人才培养资助专项科研基金资助项目(2009 D005001000005)

通讯作者:

作者简介: 蒋效宇(1979—), 男, 副教授、博士, 主研方向: 人工智能, 自动文摘

通讯作者E-mail: sxyjiangxiaoyu@bift.edu.cn

参考文献:

- [1] 傅问莲, 陈群秀. 基于规则和统计的中文自动文摘系统[J]. 中文信息学报. 2006, 20(5): 10-16 
- [4] 张虹. 基于自动文本分类的关键词抽取算法[J]. 计算机工程. 2009, 35(12): 145-147 [浏览](#)
- [7] 尤文建, 李绍滋, 李堂秋. 基于词汇链的文本过滤模型[J]. 计算机应用研究. 2003, 20(9): 32-35 

本刊中的类似文章

1. 向河林, 张明西, 李珀瀚, 何震瀛, 汪卫. 一种基于聚类的语义检索算法[J]. 计算机工程, 2012, 38(2): 36-38
2. 钟将, 孙启干, 李静. 基于归一化向量的文本分类算法[J]. 计算机工程, 2011, 37(8): 47-49
3. 李庆诚, 左珊珊, 董振华, 张金. 中文RSS信息自动检索与分类研究[J]. 计算机工程, 2011, 37(6): 79-81

扩展功能

本文信息

Supporting info

PDF(330KB)

[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

自动文摘

向量空间模型

关键词抽取

词汇链

未登录词识别

本文作者相关文章

蒋效宇

PubMed

Article by Jiang, X. Y.

4. 蒋凯, 关侏红. 基于重启型随机游走模型的图上关键字搜索[J]. 计算机工程, 2011,37(3): 42-43,46
5. 张荐硕, 方钰. 基于向量空间模型的Web服务发现方法[J]. 计算机工程, 2011,37(3): 36-38
6. 陈燕, 龙建勋. 基于明确语义分析的自动文摘算法[J]. 计算机工程, 2011,37(3): 183-185
7. 黄小亮, 郁抒思, 关侏红. 基于LDA主题模型的软件缺陷分派方法[J]. 计算机工程, 2011,37(21): 46-48
8. 王萌, 徐超, 李春贵, 何婷婷. 基于子主题区域划分的多文档自动文摘方法[J]. 计算机工程, 2011,37(12): 158-160
9. 台德艺; 王 俊 . 文本分类特征权重改进算法[J]. 计算机工程, 2010,36(9): 197-199,
10. 艾伟, 孙四明, 张峰. 基于本体的Web文本挖掘与信息检索[J]. 计算机工程, 2010,36(22): 75-77

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="5601"/>
5			