中 国 中 药 杂 志
China Journal of Chinese Materia Medica

正则表达式在上市中药文献信息提取中的应用

作者中文名	作者英文名	单位中文名	单位英文名	E-Mail
王志飞	WANG Zhifei	中国中医科学院 中医临床基础医学研究所, 北京 100700 中国中医科学院 博士后流动站, 北京 100700	Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China Post-doctoral Station of China Academy of Chinese Medical Sciences, Beijing 100700, China	
谢雁鸣	XIE Yanming	中国中医科学院 中医临床基础医学研究所, 北京 100700	Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China	zhinanb2010@yahoo.com.cn
王永炎	WANG yongyan	中国中医科学院 中医临床基础医学研究所, 北京 100700	Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China	

中文摘要：上市中药文献分析任务繁重,利用计算机自动提取目标信息之后再进行人工检查,可以简化工作、提高效率。该文分析了上市中药文献信息提取的难点,认为非术语信息的匹配和不同分组的同类信息的区分是影响常用信息提取技术在上市中药文献信息提取中应用的两大关键问题。该文分析了正则表达式的模式匹配对于解决上述两大问题的价值,并选取"病例数"(非术语信息)和"有效率"(不同分组同类信息区分)2种信息,以此为例探讨了基于正则表达式的信息提取方法,初步实践了这一思路。

中文关键词：中药上市后再评价 文献分析 正则表达式 信息提取

Application of regular expression in extracting key information from Chinese medicine literatures about re-evaluation of post-marketing surveillance

Abstract: Computerizing extracting information from Chinese medicine literature seems more convenient than hand searching, which could simplify searching process and improve the accuracy.However, many computerized auto-extracting methods are increasingly used, regular expression is so special that could be efficient for extracting useful information in research.This article focused on regular expression applying in extracting information from Chinese medicine literature.Two practical examples were reported in this article about regular expression to extract "case number (non-terminology)" and "efficacy rate (subgroups for related information identification)", which explored how to extract information in Chinese medicine literature by means of some special research method.

keywords:re-evaluation post-marketing surveillance Chinese medicine literature regular expression information extraction

查看全文　查看/发表评论　下载PDF阅读器