

● 刘玉琴¹, 彭茂祥²

(1. 中国科学技术信息研究所, 北京 100038; 2. 北京理工大学 管理与经济学院, 北京 100081)

国内外专利分析工具比较研究*

摘要: 介绍了专利分析概念、一般流程, 专利分析工具的主要功能; 依据分析工具可分析的数据源与功能将其分为专利采集加工工具, 文本挖掘与可视化工具, 专利检索分析工具, 计算机辅助创新工具四类, 并从工具所属国家、功能类型、分析数据源、主要功能、结果呈现 5 个方面对国内外常用的 26 种专利分析工具进行系统比较; 探讨了国内专利分析工具开发应用中面临的问题, 提出国内专利分析工具应用发展的建议。

关键词: 专利信息; 专利分析; 文本挖掘; 比较研究

Abstract: The paper describes the concept of patent analysis, its general process, and the main functions of the patent analysis tools. According to the data sources the patent analysis tools can analyze and the functions of the patent analysis tools, the paper divides them into 4 categories, namely, the patent collection and processing tools, the text mining and visualization tools, the patent retrieval and analysis tools, and the computer-aided innovation tools. The paper makes a systematic comparison of commonly-used 26 patent analysis tools at home and abroad in terms of the countries they belong to, the types of the functions, the data sources to be analyzed, the main functions, and the result presentation. Finally, the paper discusses the problems in the development and application of the patent analysis tools at home, and offers some suggestions for the application and development of patent analysis tools.

Keywords: patent information; patent analysis; text mining; comparative study

现代社会进入了信息社会, 信息资源已成为现代社会中最重要的战略资源之一。专利作为一种特殊的信息和战略资源, 在国家信息资源建设开发与利用中有着特殊的地位和作用。然而, 在面对海量的专利资源时, 如何高效、准确地开展分析工作, 对专利分析人员提出了新的挑战。可以说, 专利分析挖掘方法、挖掘工具的合理使用决定了专利获取的准确性和有效性, 以至于决定了分析人员最终的专利分析水平、效率以及专利分析质量和效益。本文重点对国内外常用的专利分析工具进行比较研究, 以期为国内开展专利分析工作和专利分析工具研发提供借鉴。

1 专利分析概述

专利分析也称专利信息分析或专利情报分析, 它是竞争情报分析的重要形式, 是在对专利文献进行筛选、鉴定、整理基础上, 利用文献计量学方法, 对其所含的各种信息要素进行统计、排序、对比、分析和研究, 从而揭示专利文献的深层动态特征, 了解技术、经济发展的历史及现状, 进行技术评价和技术预测^[1]。

专利分析过程主要分为数据检索获取、清洗加工、分析应用 3 个阶段, 见图 1。数据检索获取是专利分析的基础

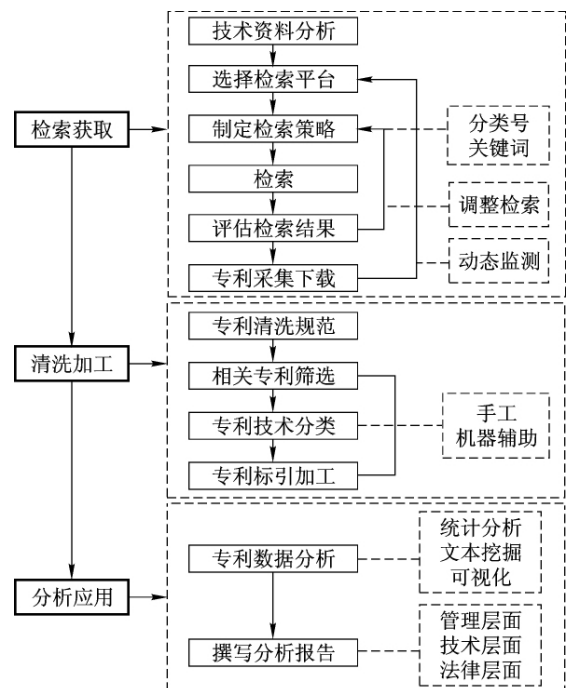


图1 专利分析的流程

* 本文为科学技术部政策法规司项目“知识产权信息服务平台建设研究”的成果, 项目编号: 02201006B。

基础性工作,从目标技术领域资料分析开始,选择检索平台,制定检索策略,试检索,评估检索结果,调整检索条件,到检索结果下载。清洗加工是为了保证分析结果的准确性而对数据进行的二次加工处理,比如申请机构、发明人名称规范,相关专利筛选,技术分类,专利的技术性、创新性、风险性标注。清洗加工的手段一般采用人工的方式进行,随着文本挖掘技术的应用,有些分析工具和软件平台也加入了一些计算机辅助的手段,用以减轻人工的工作量。分析应用则是专利数据和专利分析价值的体现,分析的方法与应用的目的紧密相关,从基本的维度统计到文本挖掘、信息可视化技术的应用,为用户提供管理决策、技术研发、法律诉讼等多个层次的支撑与服务。

2 专利分析工具的主要功能

随着信息技术的飞速发展,众多专利分析工具应运而生。尽管不同分析工具各有专长,但总体来说,其功能可归纳为专利数据监测、数据采集、数据清洗、数据加工、统计分析、文本挖掘、信息可视化 7 项。

2.1 数据监测

数据监测是指用户在某些专利检索平台上进行专利检索后,将检索条件保留在检索平台或本地检索管理工具,只要保留的检索条件返回的检索结果有变化,检索平台或本地检索管理工具就会将变化信息以邮件的形式通知用户,提醒用户其所关注的专利数据发生了变化。

2.2 数据采集

一些本地化的专利分析工具中,工具本身和工具提供商并没有专利数据,只是提供指向各国官方专利局专利检索平台的数据采集功能,在用户进行检索后将检索的专利数据批量采集到本地,然后在本地进行数据加工和分析。

2.3 数据清洗

专利数据清洗一方面是对原始专利数据进行规范化操作,如申请机构和发明人名称规范、引证信息规范;另一方面允许用户对检索结果相关性进行判断,筛选出符合条件的专利集合,从而为专利分析提供准确的数据基础。

2.4 数据加工

数据加工也称数据标引,是指通过人工解读专利数据后,将专利按照预先定义的技术类别进行分类,同时,对专利所解决的技术问题、采取的技术手段、达到的技术效果、创新程度等进行人工标注,从而提炼出隐含于专利中的更加明确的技术信息。

2.5 统计分析

统计分析是指依据专利的著录项目,对专利申请时间、申请人、申请机构、申请国家、同族专利数量、引证专利数量等指标进行组合统计,用于把握专利的分布状况

及其发展态势。

2.6 文本挖掘

专利文本挖掘本质是将文本挖掘技术应用于专利文本的过程。从理论上说,任何文本挖掘技术都可在专利中进行应用,而现实应用中以专利的自动分类、自动标引、主题聚类、主题关联、机器翻译居多。一些高级的专利分析工具常把文本挖掘与可视化结合,形成技术图谱,在揭示技术领域分布和技术发展趋势方面具有广泛的应用。

2.7 信息可视化

应用于专利的信息可视化方法主要有基于社会网络分析法的网络图,如科研合作网络、引证网络、共词网络、关联网络,以及和文本挖掘密切结合的技术主题图^[2]和技术热力图^[3],用以揭示技术领域分布。

3 国内外常用专利分析工具比较

国内外较为成熟且商品化的专利分析工具有几十种,分别实现了专利数据监测、采集、清洗、加工标引、统计、文本挖掘、信息可视化,主要的专利分析工具功能和特点见表 1。依据这些工具可分析的数据源与分析功能,本文将其分为以下 4 类。

3.1 专利采集加工工具

目前,在我国范围内应用最为广泛的专利采集加工工具,以北京理工大学知识发现与数据分析实验室的 Itg-Mining^[4],保定大为 PatentEx^[5],恒河顿 HIT_恒库^[6],北京彼速 BizSolution^[7],台湾连颖科技 PatentGuider^[8]、汉之光华专利情报分析系统^[9]为代表。其主要特点在于:

1) 通过网络爬虫对各国专利局的专利进行采集、清洗、入库,提供专利的二次加工,如筛选、分类、标引等,可用以辅助进行深入的专利技术分析。这类软件可以从各个国家官方的网站获得最新最全的专利数据,但是每个系统的单个专题数据库最大容量、二次数据加工的能力差别较大。

2) 分析数据源为采集专利的结构化部分,可从时间、地域、申请机构、发明人、法律状态等角度进行不同维度的组合分析。除 ItgMining 具有少量的文本挖掘、可视化功能外,其他软件均缺少文本挖掘、信息可视化等深入的分析技术和手段,报表以简单的条形、柱形、饼形、线形图为主,少数有简单的专利引证图。

3) 价格较低,一般一次性收费 3 万~8 万不等。使用的用户数量大多有所限制,增加用户数量需按照使用的计算机数量增加使用费用,但增加额度不高。

4) 适合我国现阶段知识产权的发展需求,在国内用户群体广泛,多为知识产权中介服务机构和少数中小企业。

表 1 国内外 26 种专利分析工具比较

序号	产品名称	所属国家、机构	功能类型	分析数据源	主要功能								结果呈现	
					数据监测	数据采集	数据清洗	数据加工	基本统计	文本挖掘与可视化分析			可视化呈现	自动报告
										共现分析	聚类分析	引证分析		
1	ItgMining	中国 北京理工大学	专利采集加工	结构化、非结构化专利数据	有	有	有	有	有	有	无	有	列表、趋势图、柱形图、网络图	有
2	PatentEx	中国 保定大为	专利采集加工	结构化专利数据	无	有	有	有	有	无	无	有	列表、趋势图、柱形图	无
3	HIT_恒库	中国 恒河顿	专利采集加工	结构化专利数据	无	有	有	有	有	无	无	有	列表、趋势图、柱形图	有
4	BizSolution	中国 北京彼速	专利采集加工	结构化专利数据	无	有	有	有	有	无	无	有	列表、趋势图、柱形图	无
5	PatentGuider	中国 台湾连颖科技	专利采集加工	结构化专利数据	有	有	有	有	有	无	无	有	列表、趋势图、柱形图	无
6	汉之光华专利情报分析系统	中国 汉之光华	专利采集加工	结构化专利数据	无	有	有	有	有	无	无	无	列表、趋势图、柱形图	无
7	ItgInsight	中国 北京理工大学	文本挖掘可视化	结构化、非结构化数据	无	无	无	无	有	有	有	有	网络图	有
8	True-Teller	日本 野村研究所	文本挖掘可视化	非结构化数据	无	无	有	无	无	有	有	无	热力图、网络图	有
9	Vantage-Point	美国 GIT 技术政策与评估中心	文本挖掘可视化	非结构化数据	无	无	有	有	有	有	有	无	矩阵、网络图	有
10	ClearForest	美国 汤姆森·路透	文本挖掘可视化	结构化、非结构化数据	无	无	有	无	无	有	有	无	列表、矩阵、聚类图	无
11	OminiViz	英国 Biowisdom	文本挖掘可视化	结构化、非结构化数据	无	无	有	无	有	有	有	无	星系图、主题图	无
12	RefViz	美国 汤姆森·路透	文本挖掘可视化	结构化数据	无	无	有	无	有	有	有	无	星系图和矩阵	无
13	STN AnaVist	美国 化学协会	文本挖掘可视化	结构化、非结构化数据	无	无	有	无	有	有	有	无	列表、景观图	无
14	Thomson Data Analyzer	美国 汤姆森·路透	文本挖掘可视化	结构化、非结构化数据	无	无	有	有	有	有	有	无	列表、图表、矩阵、聚类图	有
15	Vxinsight	美国 Sandia 国家实验室	文本挖掘可视化	结构化、非结构化数据	无	无	无	无	无	无	有	无	聚类图	无
16	Aureka	美国 汤姆森·路透	文本挖掘可视化	结构化、非结构化专利数据	有	无	有	无	有	有	有	有	主题图、引文树、聚类图	有
17	东方灵盾中外专利检索及战略分析平台	中国 东方灵盾	专利检索分析	结构化、非结构化专利数据	无	无	有	有	有	无	有	有	列表、趋势图、柱形图、聚类图	有
18	Wisdomain	美国 Wisdomain	专利检索分析	结构化、非结构化专利数据	无	无	有	无	有	有	有	有	列表、图表、树图、引文图	无
19	Delphion	美国 汤姆森·路透	专利检索分析	结构化、非结构化专利数据	有	无	有	无	有	无	有	有	列表、引文树、聚类图	无
20	Innovation	美国 汤姆森·路透	专利检索分析	结构化、非结构化专利数据	有	无	有	有	有	有	有	有	聚类图、引证图、网络图	无
21	Innography	美国 Dialog	专利检索分析	结构化、非结构化专利	有	无	有	无	无	无	无	无	气泡图、组合图	无
22	TotalPatent	美国 LexisNexis	专利检索分析	结构化、非结构化专利数据	有	无	有	无	无	无	无	有	表、趋势图、柱形图	无
23	QUESTEL ORBIT	法国 电信多媒体	专利检索分析	结构化、非结构化专利数据	有	无	有	无	有	有	有	有	聚类图、引证图、网络图	无
24	WIPS	韩国 世界知识产权检索株式会社	专利检索分析	结构化专利数据	有	无	有	无	有	无	无	有	列表、趋势图、饼图、柱形图	无
25	Goldfire Innovator	美国 Invention Machine	计算机辅助创新	非结构化专利数据	无	无	有	无	有	无	无	有	分类图、趋势图	无
26	Pro/Innovator	中国 亿维讯	计算机辅助创新	非结构化专利数据	无	无	有	无	有	无	无	无	树形图、矩阵	无

3.2 文本挖掘与可视化工具

这类工具主要来源于欧、美、日等国，在我国高校和

科研院所内应用广泛，以日本野村研究所的 True-Teller^[10]，英国 Biowisdom 公司的 OminiViz^[11]，美国汤姆森·

路透的 ClearForest、RefViz、Thomson Data Analyzer^[12]，化学协会的 STN AnaVist^[13]，Sandia 国家实验室的 Vxin-sight^[14] 为代表。国内同类工具有北京理工大学知识发现与数据分析实验室的科技文本可视化挖掘系统 ItgIn-sight^[15]，相比同类国外软件其用户群体非常有限。这类软件工具的主要特点在于：

1) 分析数据源为结构化、非结构化文本数据。结构化数据包括时间、地域、机构、作者等，可从不同维度进行组合分析。非结构化数据主要是指大段的文本数据，如专利的摘要、专利原文等，采用自然语言理解、文本挖掘、信息可视化技术进行非结构化信息的分析与结果呈现。

2) 分析的技术先进，融合了文本挖掘、信息可视化等技术；表现形式规整美观，以网络图、聚类图为主。分析对象不仅限于专利数据，任何文本数据，只要符合软件的输入标准，即可进行分析。

3) 价格普遍较高，一般按年收费，价格从几万到上百万不等，并有一部分产品由于美国的对华技术出口限制，禁止在我国境内销售，如日本野村研究所 True-Teller，只能提供分析服务，不能提供软件产品，而一次分析服务的价格在 10 万人民币以上。

4) 用户群多为高校与科研院所的研究人员，且以科技论文分析居多，在专利分析领域的应用并不广泛。

5) 国内这类软件工具，特别是产品化相对成熟的非常少。

3.3 专利检索分析工具

这类工具主要在欧、美、日以及亚洲国家范围内应用，产品提供者以欧美为主。典型代表有汤姆森·路透的 Innovaton^[16]、Aureka^[17]、Delphion^[18]，美国 LexisNexis 的 TotalPatent^[19]，法国电信多媒体公司的 Questel Orbit^[20]，韩国世界知识产权检索株式会社的 WIPS^[21]。其主要特点在于：

1) 集成海量专利数据资源，并对这些数据进行了规范，用户登录平台可以检索世界范围内绝大多数国家的专利，而且由于部分数据经过加工，增强了检索的准确性。

2) 拥有先进的数据检索、文本挖掘、信息可视化技术，以数据检索为主要功能，数据分析为辅。

3) 缺乏根据用户需求对数据进行二次加工的功能，不能存储用户的数据，使用期限到时，用户数据无法保留，增加数据管理的风险。同时，这类工具价格高昂，尤其是欧美的产品，一般按年收取服务费，数据越丰富、检索功能越强大，价格越高。

4) 用户以高校、科研院所和拥有较强研发能力的大型企业为主。

5) 国内除国家知识产权局拥有官方的专利检索平台外，还存在少量商业化的检索分析工具，但与欧美同类工具相比，国内该类工具数据基础、检索分析技术和用户普及程度差距较大。

3.4 计算机辅助创新工具

在国内应用比较广泛的计算机辅助创新工具有美国 Invention Machine 公司的 Goldfire Innovator^[22] 和中国北京亿维讯科技有限公司的 Pro/Innovator^[23]。其主要特点在于：

1) 基于发明问题解决理论选取解决典型技术问题的专利进行加工，提炼出解决问题的方法。采用类比方式将该方法应用到其他类似技术问题中，即实践、理论、再实践的转化过程。

2) 以解决微观技术问题为目的，在管理、决策支持方面应用有限。

4 国内专利分析工具开发应用中面临的问题探讨

4.1 国外相关技术与产品的对华出口限制问题

尽管欧美已在专利深度挖掘方面取得了一定的进展，并开发设计了一些相关的软件工具，然而出于种种原因，其中许多工具属于美国限制出口的技术和软件，尚不让我我国进口，如美国 Sandia 国家实验室的 VxInsight 被禁止在华销售^[24]；应用了美国核心文本挖掘技术的日本专利组合分析工具 True-Teller，被限制在美国盟国范围内使用。这从一个侧面反映了专利智能挖掘技术的重要性，同时，也应认识到开展我国自主的挖掘算法和挖掘工具研究是必要的。

4.2 国外相关技术的知识产权壁垒问题

在海量信息挖掘与可视化方面，欧美企业与研发机构为了自身的商业利益，将其研究成果申报专利，用以束缚其他企业和机构使用同类技术。如在引证分析上，美国 INXIGHT 软件公司的专利 US 6300957，将引证关系双曲树可视化表示申请了专利。在语义分析上，IBM 公司的专利 US 6006223 将基于语义的专利技术趋势挖掘方法申请了专利。这种以专利保护海量信息挖掘与可视化技术及其在科技情报分析领域中应用的趋势在日本、韩国也有所显露。随着国外企业对华市场重视程度的加强，其知识产权保护保护的触角已经开始向国内延伸。

4.3 国外相关服务的价格问题

在信息服务方面，国外一些信息情报服务企业凭借其丰富的数据资源和领先的行业经验，在为国内相关企业提供信息服务时，往往提出较高的价格。如汤姆森的 Innovation 单个账户的每年使用费为十几万人民币，Dialog 的 Innography 单个账户的每年使用费也在 10 万人民币以上，True-Teller 的一次分析服务就高达 10 万人民币。开展我国

自主的挖掘算法和分析工具对于有效抑制国外相关服务价格具有重要意义。

4.4 国内产品技术水平不高的问题

国内一些软件企业研发和设计了各种专利分析工具,如大为软件的 PatentEx、东方灵盾的专利检索与分析平台。这些软件平台以数据采集、数据管理见长,深入的文本挖掘与可视化技术基本没有,数据分析的能力较弱,并且缺少完备的专利数据支撑,使得其产品同国外同类产品的技术水平相差较大,销售范围有限,有些甚至没有在市场上销售。

5 国内专利分析工具应用发展的建议

1) 基于开源软件进行高水平专利分析工具的自主研发。在有效规避知识产权问题的条件下,利用开源软件的技术优势和技术开放特征,进行融合文本挖掘和信息可视化技术的高水平专利分析工具的研发,进而缩短研发周期。

2) 促进专利信息加工和专利分析流程的标准化推广。建立专利信息二次加工标引的规范,促进专利分析的流程化、标准化,进而促进国内专利分析工具数据格式与功能模块的标准化,提高不同工具间数据的兼容性。

3) 加强科研院所与中小企业间合作。一方面将科研院所拥有的专利资源和专利分析平台一定程度地向国内中小企业开放,实现资源共享;另一方面,将企业专利分析的需求向高校或科研院所进行转接,更好地发挥科研院所的专利资源和专利分析优势。

4) 加强专利资源共享,建立专利资源与专利分析的共享服务平台与服务机制。结合国家的科技发展战略,由科技主管部门组织建设行业专利信息服务平台,共享专利资源和专利分析技术。

6 结束语

本文对国内外的专利分析工具进行了系统的研究和比较,对国内专利分析工具应用开发中面临的问题进行探讨,并提出国内专利分析工具应用发展的建议。一方面为国内开展专利分析工作提供借鉴;另一方面为国内专利分析工具研发、信息服务提供决策参考。在下一步的研究中,将继续开展国内专题专利数据库的构建和准入标准研究,以期为建设共享的行业知识产权信息服务平台提供决策支撑。□

参考文献

- [1] 王敏,李海存,许培扬. 国外专利文本挖掘可视化工具研究 [J]. 图书情报工作, 2009, 53 (24): 86-90.
- [2] WISE J A, THOMAS J J, PENNOCK K, et al. Visualizing

the non-visual: spatial analysis and interaction with information from text documents [C] //The IEEE Symposium on Information Visualization '95, Atlanta, Georgia, USA: IEEE Computer Society Press, 1995: 51-58.

- [3] MIYAKE M, YUJI M, KEIICHI H. Strategic intellectual property portfolio management: technology appraisal by using the technology heat map [EB/OL]. [2012-01-01]. <http://www.nri.co.jp/english/opinion/papers/2004/pdf/np200483.pdf>.
- [4] 依格专利工程系列软件 [EB/OL]. [2012-01-01]. <http://blog.sciencenet.cn/home.php?mod=space&uid=394038&do=blog&id=510333>.
- [5] PatentEX 专利信息创新平台 [EB/OL]. [2012-01-01]. <http://www.daweisoft.com/Production/Demo/PatentEX.htm>.
- [6] HIT_恒库 [EB/OL]. [2012-01-01]. <http://www.all-patent.com/product/hit/hit-2.html>.
- [7] BizSolution [EB/OL]. [2012-01-01]. <http://www.bizsolution.com.cn/Product/PatentSearch.aspx>.
- [8] PatentGuider [EB/OL]. [2012-01-01]. <http://www.csip.org.cn/col/ipzszx/2008/5/9/08590HDKDJK538DI7FJC1.html>.
- [9] 汉之光华专利情报分析系统 [EB/OL]. [2012-01-01]. <http://old.iprtop.com/>.
- [10] True-Teller [EB/OL]. [2012-01-01]. <http://www.true-teller.net/textmining/patent>.
- [11] BioWisdom. OminiViz [EB/OL]. [2012-01-01]. <http://www.biowisdom.com/content/omniviz>.
- [12] Thomson data analyzer [EB/OL]. [2012-01-01]. <http://science.thomsonreuters.com.cn/productservices/TDA>.
- [13] STN anavist [EB/OL]. [2012-01-01]. <http://www.cas.org/products/anavist/index.html>.
- [14] DAVIDSON G S, HENDRICKSON B, JOHNSON D K, et al. Knowledge mining with VxInsight: discovery through interaction [J]. Journal of Intelligent Information Systems, 1998, 11 (3): 259-285.
- [15] ItgInsight [EB/OL]. [2012-01-01]. <http://blog.sciencenet.cn/home.php?mod=space&uid=394038&do=blog&id=494512>.
- [16] 汤姆森开发新的知识产权分析工具 Thomson Innovation [J]. 现代图书情报技术, 2008 (5): 101.
- [17] 陈燕, 邓鹏, 李芳. AUREKA 信息平台介绍 [J]. 中国发明与专利, 2007 (5): 63-65.
- [18] 徐勇. Delphion 知识产权网站专利信息检索 [J]. 现代图书情报技术, 2001 (5): 46-47.
- [19] LexisNexis [EB/OL]. [2012-01-01]. <http://www.lexisnexis.com/en-us/products/total-patent.page>
- [20] QUESTEL [EB/OL]. [2012-01-01]. <http://www.questel.orbit.com>.
- [21] 赵旭, 唐恒. 中外四大专利分析软件的功能概述及综合比较 [J]. 图书情报研究, 2010, 3 (4): 50-54.
- [22] The goldfire innovator solution [EB/OL]. [2012-01-01]. <http://www.randit.com/allegati/trizsixsigma/tecnologie/brochure%20eng%20randit2.pdf>.
- [23] Pro/innovator [EB/OL]. [2012-01-01]. <http://www.iwint.com.cn/ch/index.html>.
- [24] 王柏, 吴巍, 徐超群. 复杂网络可视化研究综述 [J]. 计算机科学, 2007, 34 (4): 17-22.

作者简介: 刘玉琴, 男, 1979 年生, 高级工程师, 博士。

研究方向: 专利分析, 信息可视化。

彭茂祥, 男, 1970 年生, 博士生。研究方向: 技术预测, 专利分析。

收稿日期: 2012-03-26