

# (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2017年6月29日 (29.06.2017)



(10) 国际公布号  
WO 2017/107010 A1

- (51) 国际专利分类号:  
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/098086
- (22) 国际申请日: 2015年12月21日 (21.12.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 浙江核新同花顺网络信息股份有限公司 (HITHINK ROYALFLUSH INFORMATION NETWORK CO., LTD.) [CN/CN]; 中国浙江省杭州市文二西路1号901室, Zhejiang 310012 (CN)。
- (72) 发明人: 易峥 (YI, Zheng); 中国浙江省杭州市文二西路1号901室, Zhejiang 310012 (CN)。夏炜 (XIA, Wei); 中国浙江省杭州市文二西路1号901室, Zhejiang 310012 (CN)。陶志伟 (TAO, Zhiwei); 中国浙江省杭州市文二西路1号901室, Zhejiang 310012 (CN)。潘杭平 (PAN, Hangping); 中国浙江省杭州市文二西路1号901室, Zhejiang 310012 (CN)。
- (74) 代理人: 北京汉昊知识产权代理事务所 (普通合伙) (HANHOW INTELLECTUAL PROPERTY); 中国北京市东长安街1号东方广场西1区1111室, Beijing 100738 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,

[见续页]

(54) Title: INFORMATION ANALYSIS SYSTEM AND METHOD BASED ON EVENT REGRESSION TEST

(54) 发明名称: 基于事件回测的信息分析及方法

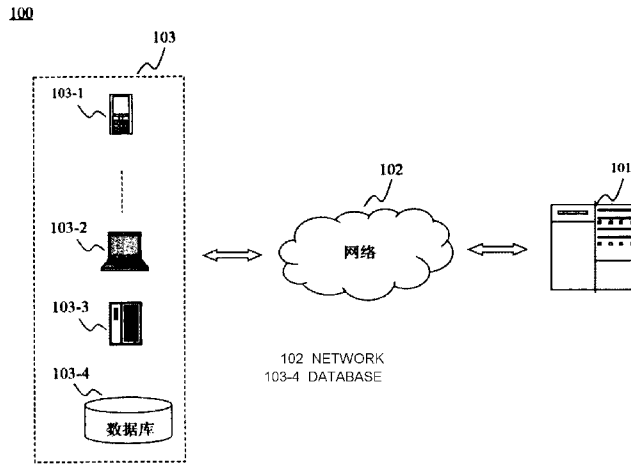


图 1

(57) Abstract: One aspect of the present invention relates to an information analysis method based on an event regression test. The information analysis method can comprise: collecting information, pre-processing the information and extracting an entity in the information, processing the information and extracting a relevant attribute in the information, determining the type of information and generating a refinement event, generating a corresponding natural language sentence according to the generated refinement event, performing a regression test on the generated natural language sentence, and generating a regression testing report, etc. In addition, the present invention relates to a user being able to select any announcement or news to perform a regression test, and generate a regression testing report. Moreover, the present invention relates to a user being able to input a natural language sentence and perform a regression test on an input natural language sentence so as to generate a regression testing report. Furthermore, the present invention relates to an information analysis system based on an event regression test, comprising a collection module, a system database, a processing module, a natural language processing module and a regression testing module.

(57) 摘要:

[见续页]



WO 2017/107010 A1



---

RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, **本国际公布:**  
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, — 包括国际检索报告(条约第 21 条(3))。  
TG)。

---

本发明一方面涉及一个基于事件回测的信息分析方法。该信息分析方法可以包括收集信息，对信息进行预处理并提取信息中的实体，处理信息并提取信息中的相关属性，判断信息类别并生成细化事件，根据生成的细化事件生成对应的自然语言语句，对生成的自然语言语句进行回测，生成回测报告等。本发明另一方面涉及用户可以选择任何公告或者新闻进行回测，并生成回测报告。本发明另一方面涉及用户可以输入自然语言语句并对输入的自然语言语句进行回测，生成回测报告。本发明另一个方面涉及一种基于事件回测的信息分析系统，包括收集模块、系统数据库、处理模块、自然语言处理模块和回测模块。

## 基于事件回测的信息分析及方法

### 技术领域

[0001] 本发明涉及一种信息分析及方法，尤其是对一个事件获得的相关信息、自然语言语句进行自动分析，从而获得其历史回测信息等。

### 背景技术

[0002] 随着互联网的不断普及，人们越来越习惯于使用互联网获取信息或根据数据分析信息。由于互联网覆盖范围的不断扩大与信息的不断增加，数据也在不断增加，当人们试图使用互联网获取某种信息时，经常会遇到信息篇幅长、信息内容多样等特点，需要占用时间去阅读和分析。同时，有的行业领域或用户需要根据当前信息回测历史数据来做出决定。例如在金融行业领域中，研究交易策略或投资策略时，使用回测可以评估该策略在过去一段时间内的表现和有效性，从而帮助投资者进行投资决策的分析。又如在天气预报中，根据实时的温度、湿度风向和气压的数据，可以通过分析同等条件下的历史天气情况以预测未来天气的情况等。

### 简述

[0003] 本发明一方面是关于一种信息分析系统，根据其中一个实施例，该信息分析系统包括一种计算机可读的存储媒介，所述存储媒介存储可执行模块，该存储媒介包括收集模块，所述收集模块能够收集信息；处理模块，所述处理模块能够对收集的信息进行预处理，从预处理后的信息中提取事件；自然语言处理模块，所述自然语言处理模块能够根据提取出的事件生成自然语言语句；回测模块，所述回测模块能够根据生成的自然语言语句结合历史信息生成回测结果。一个处理器，所述处理器能够执行所述计算机可读的存储媒介存储的可执行模块。

[0004] 根据本发明的另一个实施例，该信息分析系统进一步包括一

个数据库，所述数据库能够储存所述的收集信息、预处理后的信息、提取的事件、自然语言语句、历史信息、回测结果。

5 [0005] 根据本发明的另一个实施例，该数据库包括原始信息数据库、文本数据库、文本预处理数据库、实体数据库、事件属性数据库、关键词数据库、文本分类数据库、历史信息数据库、自然语言处理数据库、事件识别数据库、回测模块数据库、文字模板数据库、词典数据库。

10 [0006] 根据本发明的另一个实施例，该信息分析系统的处理模块进一步包括格式转换模块、文本处理模块、属性抽取模块、事件识别模块。

[0007] 根据本发明的另一个实施例，该处理模块进一步包括文本分类模块。

15 [0008] 根据本发明的另一个实施例，该处理模块采用的方法包括卡方统计、信息增益、互信息、几率比、交叉熵、类间信息差、关键词统计、决策树、Rocchio、朴素贝叶斯、神经网络、支持向量机、线性最小平方拟合、最邻近算法、遗传算法、情感分类、最大熵、Generalized Instance Set、同义词配置、布尔关联规则、位置规则、机器学习。

20 [0009] 根据本发明的另一个实施例，该自然语言处理模块可以从收集模块接收信息。

[0010] 根据本发明的另一个实施例，该回测模块进一步包括回测信息判断，所述回测信息判断根据回测结果的情况给出评价。

[0011] 根据本发明的另一个实施例，该回测结果可以被展示给用户。

25 [0012] 本发明的另一方面涉及一种信息分析方法，该信息分析方法包括收集信息；根据所述信息提取事件；根据所述事件生成自然语言语句；对所述自然语言语句进行回测分析。

[0013] 根据本发明的另一个实施例，该收集信息包括用户输入信息和非用户输入信息，所述非用户输入信息来源包括通信终端和服务器。

[0014] 根据本发明的另一个实施例，该收集信息包括公告信息和新

闻信息。

[0015] 根据本发明的另一个实施例，该提取事件进一步包括实体识别和属性抽取。

5 [0016] 根据本发明的另一个实施例，该实体识别进一步包括格式转换、文本分词、数字及单位归一化处理。

[0017] 根据本发明的另一个实施例，该属性抽取可以通过系统定义的模型实现。

[0018] 根据本发明的另一个实施例，该自然语言语句可以根据所述的提取事件生成。

10 [0019] 根据本发明的另一个实施例，该自然语言语句可以根据用户输入信息生成。

[0020] 根据本发明的另一个实施例，该自然语言语句进一步根据事件类别扩展。

15 [0021] 根据本发明的另一个实施例，该自然语言语句分析包括对所述自然语言语句进行回测。

[0022] 根据本发明的另一个实施例，该自然语言语句回测可以根据信息类别生成回测结果。

### 附图描述

20 [0023] 为了更清楚地说明本发明实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单的介绍。显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域的普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图将本发明应用于其它类似情景。除非从语言环境中显而易见或另做说明，图中相同标号代表相同结构和操作。

25 [0024] 图 1 所示的是信息分析系统的一种示例系统配置的示意图；

[0025] 图 2 所示的是信息分析系统的模块示意图；

[0026] 图 3 所示的是信息分析流程图；

[0027] 图 4 所示的是收集模块的结构示意图；

[0028] 图 5 所示的是处理模块的结构示意图；

- [0029] 图 6 所示的是格式转换模块的结构示意图；
- [0030] 图 7 所示的是文本预处理模块的结构示意图；
- [0031] 图 8 所示的是文本分类模块的结构示意图；
- [0032] 图 9 所示的是属性抽取模块的结构示意图；
- 5 [0033] 图 10 所示的是处理模块的流程图；
- [0034] 图 11 所示的是自然语言处理模块的结构示意图；
- [0035] 图 12 所示的是回测模块的结构示意图；
- [0036] 图 13 所示的是回测流程图；
- [0037] 图 14 所示的是系统数据库的结构示意图；
- 10 [0038] 图 15 所示的是信息分析流程图；
- [0039] 图 16 所示的是信息分析系统线上工作流程图；
- [0040] 图 17 所示的是信息分析系统针对新闻或公告的一个交互界面示意图；
- [0041] 图 18 所示的是信息分析系统针对用户输入的一个交互界面示意图；
- 15 [0042] 图 19 所示的是信息分析系统所用到的公告文本图。

### 具体描述

[0043] 如本说明书和权利要求书中所示，除非上下文明确提示例外情形，“一”、“一个”、“一种”和/或“该”等词并非特指单数，也可  
20 包括复数。一般说来，术语“包括”与“包含”仅提示包括已明确标识的步骤和元素，而这些步骤和元素不构成一个排它性的罗列，方法或者设备也可能包含其它的步骤或元素。

[0044] 本说明书所述的信息分析方法是指通过收集信息，处理信息，生成自然语言语句，分析数据提供参考信息等。在一些实施例中，本  
25 发明一方面涉及一种信息分析系统。该信息分析系统可以包括收集模块、系统数据库、处理模块、自然语言处理模块和回测模块。本发明另一方面涉及一个基于事件回测的信息分析方法。该信息分析方法可以包括收集信息，对信息进行预处理并提取信息中的实体，处理信息并提取信息中的相关属性，判断信息类别并生成细化事件，根据生成

的细化事件生成对应的自然语言语句，对生成的自然语言语句进行回测，生成回测报告等。本发明另一方面涉及用户可以选择任何公告或者新闻进行实时回测，并生成回测报告。本发明另一方面涉及用户可以输入任何自然语言语句并对输入的自然语言语句进行回测，实时地生成回测报告。

**[0045]** 本发明的不同实施例可适用于多种领域，包括但不限于金融及其衍生物投资（包括但不限于股票、债券、黄金、纸黄金、白银、外汇、贵金属、期货、货币基金等）、科技（包括但不限于数学、物理、化学及化学工程、生物及生物工程、电子工程、通信系统、互联网、物联网等）、政治（包括但不限于政治人物、政治事件、国家）、新闻（从区域而言，包括但不限于地区新闻、国内新闻、国际新闻；从新闻主体而言，包括但不限于政治新闻、体育新闻、科技新闻、经济新闻、生活新闻、气象新闻等）等。依据本发明的至少一个实施例，可以对各种信息资源，如文字、图片、音频、视频的内容进行快速的提炼，并根据相关历史信息，提炼回测策略并生成回测报告，使用户更加方便快捷地了解信息可能的未来影响。本发明的不同实施例应用场景包括但不限于网页、浏览器插件、客户端、定制系统、企业内部分析系统、人工智能机器人等一种或多种组合。以上对适用领域的描述仅仅是具体的示例，不应被视为是唯一可行的实施方案。显然，对于本领域的专业人员来说，在了解一种基于事件回测的信息分析方法和系统的基本原理后，可能在不背离这一原理的情况下，对实施上述方法和系统的应用领域形式和细节上的各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。例如，在本发明的一个实施例中，回测报告是以统一的文字形式显示给用户的，对于本领域的专业人员来说，回测报告也可以是以统一的音频格式或视频格式显示给用户。与此类似的替换或修正或改变，仍在本发明的保护范围之内。为了更清楚地说明本发明实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单的介绍。显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域的普通技术人员来讲，在不付出创造

性劳动的前提下，还可以根据这些附图将本发明应用于其它类似情景。除非从语言环境中显而易见或另做说明，图中相同标号代表相同结构和操作。

**[0046]** 图 1 所示的是信息分析系统的一种示例系统配置的示意图。

5 示例系统配置 100 可以包含但不限于一个或多个信息分析系统 101、一个或多个网络 102 和一个或多个信息源 103。信息分析系统 101 可以用于对收集的信息进行分析加工以生成分析结果的系统。信息分析系统 101 可以是一个服务器，也可以是一个服务器群组。一个服务器群组可以是集中式的，例如数据中心。一个服务器群组也可以是分布式的，例如一个分布式系统。信息分析系统 101 可以是本地的，也可以是远程的。网络 102 可以提供信息交换的渠道。网络 102 可以是单一网络，也可以是多种网络组合的。网络 102 可以包括但不限于局域网、广域网、公用网络、专用网络、无线局域网、虚拟网络、都市城域网、公用开关电话网络等一种或多种组合。网络 102 可以包括多种网络接入点，如有线或无线接入点、基站或网络交换点，通过以上接入点使数据源连接网络 102 并通过网络发送信息。信息源 103 可以提供各种信息。信息源 103 可以包括但不限于服务器、通信终端。进一步地，服务器可以是 web 服务器、文件服务器、数据库服务器、FTP 服务器、应用程序服务器、代理服务器器等，或者上述服务器的任意组合。通信终端可以是手机、个人电脑、可穿戴设备、平板电脑、智能电视等，或则上述通信终端的任意组合。信息源 103 可以通过网络 102 发送或/和收集信息到信息分析系统 101，信息源 103 可以是用户输入的信息，也可以是其他数据库或信息源提供的信息。

**[0047]** 图 2 所示的是信息分析系统的模块示意图。信息分析系统 101 可以包含但不限于一个或多个收集模块 201、一个或多个处理模块 202、一个或多个自然语言处理模块 203、一个或多个回测模块 204、一个或多个系统数据库 205。上述的模块中部分或全部可以与网络 102 连接。上述模块可以是集中式的也可以是分布式的。上述模块中的一个或多个模块可以是本地的也可以是远程的。收集模块 201 可以主要



用于以各种方式收集所需要的信息，收集信息的方式可以是直接的（例如直接通过网络 102 从一个或多个信息源 103 获取信息）也可以是间接的（例如通过处理模块 201、自然语言处理模块 202、回测模块 204 或者系统数据库 205 来获取信息）。处理模块 202 可以主要用于信息的预处理，信息的预处理可以是人工的，也可以是自动的，信息的预处理可以包括但不限于格式转换、分词处理、实体识别、数字及单位归一化处理、文本分类、事件属性抽取、细化事件识别、加密文档解密等一种或多种组合。自然语言处理模块 203 可以主要用于生成自然语言语句，也可以接收输入的自然语言语句。自然语言处理模块 203 处理信息的方式可以是人工的，也可以是自动的。回测模块 204 可以主要用于分析信息。分析方法可以包括但不限于系统定义、用户自定义选择、机器学习等其中的一种或多种组合。信息分析可以是人工实现的，也可以是自动完成的，或者是两者相结合完成的。系统数据库 205 可以泛指具有存储功能的设备。系统数据库 205 主要用于存储从信息源 103 收集的数据和信息分析系统 101 工作中产生的各种数据。系统数据库 205 可以是本地的，也可以是远程的。系统数据库与系统其他模块间的连接或通信可以是有线的，也可以是无线的。

**[0048]** 收集模块 201 可以将收集到的信息传输给处理模块 202。收集模块 201 也可以将收集到的信息传输给自然语言处理模块 203。收集模块 201 也可以将收集到的信息传输给回测模块 204。收集模块 201 可以接收处理模块 202 发送的请求，也可以按照该请求访问系统数据库 205，以获取需要的数据。需要的数据被获取之后，收集模块 201 可以将该数据传输给处理模块 202。收集模块 201 可以接收自然语言处理模块 203 发送的请求，也可以按照该请求访问系统数据库 205，以获取需要的数据。需要的数据被获取之后，收集模块 201 可以将该数据传输给自然语言处理模块 203。收集模块 201 可以接收回测模块 204 发送的请求，也可以按照该请求访问系统数据库 205，以获取需要的数据。需要的数据被获取之后，收集模块 201 可以将该数据传输给回测模块 204。

**[0049]** 收集模块 201 可以主要用于以各种方式收集所需要的信息。收集模块 201 可以通过向信息源 103 发送请求，以获取需要的信息。收集模块 201 在获取需要的信息后，可以将所获得的信息进行下一步处理或者存储在系统数据库 205 中。收集模块 201 也可以通过向系统数据库 205 发送请求，以获取存储在系统数据库 205 中的信息。可选择地，系统数据库 205 也可以直接向信息源 103 发送请求，获取的信息可以被存储在系统数据库 205 中。信息源 103 可以是服务器、通信终端等。进一步地，服务器可以是 web 服务器、文件服务器、数据库服务器、FTP 服务器、应用程序服务器、代理服务器等，或者上述服务器的任意组合。通信终端可以是手机、个人电脑、可穿戴设备、平板电脑、智能电视等，或者上述通信终端的各种组合。上述需要的信息可以包括但不限于各种新闻、研报、公告、消息、报告、通知、论文、期刊等中的一种或多种。上述需要的信息可以是关于各个行业的信息，包括但不限于体育、娱乐、经济、政治、军事、文化、艺术、科学、工程等中的一种或多种。上述需要的信息的形式可以包括但不限于文字、图片、音频、视频等中的一种或多种。例如某视频网站播放的视频新闻《世界银行下调今年全球经济增长预期至 2.8%》、某新闻网站报道的网页新闻《汇丰 5 月份中国服务业 PMI 升至 53.5》、某证券交易所发布的上市公司公告《A 股份有限公司关于签署日常经营重大合同的公告》、某体育赛事直播平台发布的足球赛事预告《本周六切尔西俱乐部将在主场斯坦福桥球场迎战同城死敌阿森纳》。

**[0050]** 处理模块 202 可以与收集模块 201 进行双向通信。处理模块 202 可以处理收集模块 201 传输的信息，信息处理可以包括但不限于格式转换、文本预处理、文本分类、属性抽取和事件识别等中的一种或多种组合。处理模块 202 也可以向收集模块 201 发送信息，发送的信息可以包括但不限于经过处理的信息以及控制信息，该控制信息可以包括但不限于信息收集方式的控制信息、信息收集时间的控制信息、信息收集来源的控制信息等。处理模块 202 可以与自然语言处理模块 203 进行双向通信。处理模块 202 可以将经过处理后的信息传输给自

然语言处理模块 203,也可以接收自然语言处理模块 203 发送的信息。处理模块 202 可以与回测模块 204 进行双向通信。处理模块 202 可以将经过处理后的信息传输给回测模块 204,也可以接收回测模块 204 发送的信息。处理模块 202 可以与系统数据库 205 进行双向通信。处理模块 202 可以将经过处理后的信息传输给系统数据库 205 进行储存,也可以在处理过程中向系统数据库 205 发送请求信息并接收系统数据库 205 发送的信息。

**[0051]** 自然语言处理模块 203 可以向收集模块 201 发送请求,收集模块 201 可以根据请求访问系统数据库 205 或从一个或多个信息源 103,以获取需要的信息。需要的信息被获取之后,收集模块 201 将该信息传输给自然语言处理模块 203。可选择地,收集模块 201 在收到从自然语言处理模块 203 发来的请求之后,也可以将收集模块 201 中的信息传输给自然语言处理模块 203,所述信息可以来自信息源 103 或系统数据库 205。自然语言处理模块 203 可以向处理模块 202 发送请求,处理模块 202 可以根据请求访问系统数据库 205,以获取需要的信息。需要的信息被获取之后,处理模块 202 将该信息传输给自然语言处理模块 203。可选择地,处理模块 202 在收到从自然语言处理模块 203 发来的请求之后,也可以将处理模块 202 中的信息传输给自然语言处理模块 203。可选择地,自然语言处理模块 203 可以直接访问系统数据库 205,并向系统数据库 205 发送请求以获取需要的信息,该信息可以被传输给自然语言处理模块 203。可选择地,系统数据库 205 可以在没有收到请求的情况下向自然语言处理模块 203 发送信息。在本发明的一种实施例中,自然语言处理模块 203 可以直接接收来自信息源 103 的自然语言语句(图中未展示),该自然语言语句可以是用户使用输入设备输入的,输入设备包括但不限于键盘、鼠标、摄像头、扫描仪、手写输入板、语音输入装置等一种或多种组合。

**[0052]** 自然语言处理模块 203 的输入信息可以是字母、数字、字符、词语、短语、语句、段落、篇章等,或其中的一种或多种,或由任意数量的标识符集合,该标识符集合可以包含一种或多种语义。可选择

地，自然语言处理模块 203 的输入信息可以是自定义的信息类型。在本发明的一些实施例中，自然语言处理模块 203 的输入信息可以被表征为一个多元组。例如：自然语言处理模块 203 的输入信息可以被表征为一个四元组 {k, c, u, d}。其中，参数 k 可以被配置为表示信息来源，信息来源可以包括但不限于收集模块 201、处理模块 202、系统数据库 205、信息源 103（图中未展示），或者上述信息来源的任意组合。参数 c 可以被配置为表示通信时间。例如：参数 c 可以被配置为表示年份、月份、日期等。通过赋予参数 c 以特定的数值，由参数 c 指定的特定时间的信息将被输入自然语言处理模块 203。参数 u 可以被配置为表示要使用的用户模型。用户模型为依据不同的用户需求而具有不同功能的数据处理模型。参数 u 在缺省的情况下表示不应用任何数据模型。参数 d 可以被配置为表示已产生信息。已产生信息指用户在自然语言处理的过程中已经生成的各种实体和属性等，各种实体和属性将在后续的自然语言处理的过程中被运用。

15 **[0053]** 自然语言处理模块 203 可以将收集到的信息进行处理，以生成自然语言语句。生成的自然语言语句可以被传输至回测模块 204 以进行回测。具体地，自然语言处理模块 203 可以向回测模块 204 发送回测请求，该请求被获准之后，自然语言处理模块 203 将生成的自然语言语句输入到回测模块 204 进行回测。可选择地，自然语言处理模块 203 也可以不发送回测请求，而是直接将生成的自然语言语句输入到回测模块 204 进行回测。在本发明的一个实施例中，回测模块 204 在接收到自然语言处理模块 203 输入的自然语言语句后，将自然语言语句进一步处理生成数据库标准访问指令，从而访问或调取相应数据库中存储的历史数据。

25 **[0054]** 在本发明的另一种实施例中，自然语言处理模块 203 可以接收处理模块 202 生成的事件，自然语言处理模块 203 可以将接收到的自然语言语句（事件）进行拼装，也可以根据回测的需要，加上额外条件。比如：针对个股事件，则需要加上“个股代码或简称”；针对行业事件回测，则需要加上“个股对应行业”；针对全市场事件（如

央行降息) 回测则不需要加任何语句。

**[0055]** 需要注意的是, 上述对自然语言处理模块 203 输入信息的描述只是为了便于理解发明, 不应被视为是本发明唯一可行的实施方案。对于本领域的专业人员来说, 在了解所需要的信息的基本原理后, 可能在不背离这一原理的情况下, 对所需要的信息的内容进行各种修正和改变, 但是这些修正和改变仍在以上描述的范围之内。比如, 自然语言处理模块 203 的输入信息可以被表征为二元组、三元组、五元组、六元组、N 元组等, 或上述信息类型的任意组合。

**[0056]** 回测模块 204 可以向收集模块 201 发送回测条件请求, 收集模块 201 可以根据请求访问系统数据库 205 获取需要的信息。需要的信息被获取之后, 收集模块 201 将该信息传输给回测模块 204。可选择地, 收集模块 201 在收到从回测模块 204 发来的请求之后, 也可以将存储在收集模块 201 中的信息传输给回测模块 204。回测模块 204 可以向处理模块 202 发送请求, 处理模块 202 可以根据请求访问系统数据库 205 获取需要的信息。需要的信息被获取之后, 处理模块 202 可以将该信息传输给回测模块 204。可选择地, 处理模块 202 在收到从回测模块 204 发来的请求之后, 也可以将存储在处理模块 202 中的信息传输给回测模块 204。回测模块 204 可以向自然语言处理模块 203 发送请求, 自然语言处理模块 203 可以根据请求访问系统数据库 205 获取需要的信息。需要的信息被获取之后, 自然语言处理模块 203 可以将该信息传输给回测模块 204。可选择地, 自然语言处理模块 203 在收到从回测模块 204 发来的请求之后, 也可以将存储在自然语言处理模块 203 中的信息传输给回测模块 204。可选择地, 回测模块 204 可以直接访问系统数据库 205, 并向系统数据库 205 发送请求以获取需要的信息, 该信息可以被传输给回测模块 204。可选择地, 系统数据库 205 可以在没有收到请求的情况下向回测模块 204 发送信息。回测模块 204 收到的输入信息可以包括但不限于字母、数字、字符、词语、语句、段落、篇章、自然语言语句等其中的一种或多种组合。输入信息的来源可以包括但不限于收集模块 201、处理模块 202、自然

语言处理模块 203、系统数据库 205、信息源 103 等其中的一种或多种组合。

**[0057]** 系统数据库 205 或系统内的其他存储设备泛指所有可以具有读/写功能的媒介。系统数据库 205 或系统内其他存储设备可以是系统内部的，也可以是系统的外接设备。系统数据库 205 或系统内其他存储设备的连接方式可以是有线的，也可以是无线的。系统数据库 205 或系统内其他存储设备可以包括但不限于层次式数据库、网络式数据库和关系式数据库等其中一种或多种组合。系统数据库 205 或系统内其他存储设备可以将信息数字化后再以利用电、磁或光学等方式的存储设备加以存储。系统数据库 205 或系统内其他存储设备可以用来存放各种信息例如程序和数据等。系统数据库 205 或系统内其他存储设备可以是利用电能方式存储信息的设备，例如各种存储器、随机存取存储器（Random Access Memory, RAM）、只读存储器（Read Only Memory, ROM）等。系统数据库 205 或系统内其他存储设备可以是利用磁能方式存储信息的设备，例如硬盘、软盘、磁带、磁芯存储器、磁泡存储器、U 盘、闪存等。系统数据库 205 或系统内其他存储设备可以是利用光学方式存储信息的设备，例如 CD 或 DVD 等。系统数据库 205 或系统内其他存储设备可以是利用磁光方式存储信息的设备，例如磁光盘等。系统数据库 205 或系统内其他存储设备的存取方式可以是随机存储、串行访问存储、只读存储等一种或多种组合。系统数据库 205 或系统内其他存储设备可以是非永久记忆存储器，也可以是永久记忆存储器。以上提及的存储设备是列举了一些例子，该系统可以使用的存储设备并不局限于此。系统数据库 205 或系统内其他存储设备可以是本地的，也可以是远程的，也可以是云服务器上的。

**[0058]** 系统数据库 205 或系统内其他存储设备对信息的操作可以包括但不限于存储、分类、排序、筛选等一种或多种组合。系统数据库 205 或系统内其他存储设备可以与信息源 103 传递或交换信息。系统数据库 205 或系统内其他存储设备可以接收信息源 103 的信息，将其存储在系统数据库 205 或系统内其他存储设备上。根据收到的指令，

系统数据库 205 或系统内其他存储设备里面的信息可以被提取，传递给信息源 103。该指令可以是直接来自于信息源 103，也可以是来自于其他模块，如收集模块 201、处理模块 202、自然语言处理模块 203、回测模块 204 等。系统数据库 205 或系统内其他存储设备可以与收集模块 201 传递或交换信息。系统数据库 205 或系统内其他存储设备可以接收收集模块 201 收集的信息，将其存储在系统数据库 205 或系统内其他存储设备上。根据收到的指令，系统数据库 205 或系统内其他存储设备里面的信息可以被提取，传递给收集模块 201。该指令可以是直接来自于收集模块 201，也可以是来自于其他模块，如处理模块 202、自然语言处理模块 203、回测模块 204、信息源 103 等。

**[0059]** 系统数据库 205 或系统内其他存储设备可以与处理模块 202 传递或交换信息。系统数据库 205 或系统内其他存储设备可以接收处理模块 202 收集的信息，将其存储在系统数据库 205 或系统内其他存储设备上。根据收到的指令，系统数据库 205 或系统内其他存储设备里面的信息可以被提取，传递给处理模块 202。该指令可以是直接来自于处理模块 202，也可以是来自于其他模块，如收集模块 201、自然语言处理模块 203、回测模块 204、信息源 103 等。

**[0060]** 系统数据库 205 或系统内其他存储设备可以与自然语言处理模块 203 传递或交换信息。系统数据库 205 或系统内其他存储设备可以接收自然语言处理模块 203 收集的信息，将其存储在系统数据库 205 或系统内其他存储设备上。根据收到的指令，系统数据库 205 或系统内其他存储设备里面的信息可以被提取，传递给自然语言处理模块 203。该指令可以是直接来自于自然语言处理模块 203，也可以是来自于其他模块，如收集模块 201、处理模块 202、回测模块 204、信息源 103 等。

**[0061]** 系统数据库 205 发送的信息可以是直接从信息源获取的信息，也可以是经过处理分析后的信息。经过处理分析的信息，可以是经过处理模块 202 处理后储存在系统数据库 205 的信息，也可以是经过自然语言处理模块 203 处理后储存的信息。系统数据库 205 或系统内其

他存储设备与其他模块信息传递的方式可以是有线的也可以是无线的，可以是直接的也可以是间接的，可以是同时进行的也可以是顺序进行的，可以是周期的也可以是非周期的等。

**[0062]** 显然，对于本领域的专业人员来说，在了解信息分析系统及方法的原理后，可能在不背离这一原理的情况下，对各个模块进行任意组合，或者构成子系统与其它模块连接，对实施上述方法和系统的应用领域形式和细节上的各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。例如，收集模块 201、处理模块 202、自然语言处理模块 203、回测模块 204 和系统数据库 205 可以是体现在一个系统中的不同模块，也可以是一个模块实现上述的两个或两个以上模块的功能，如处理模块 202 可以收集信息并生成自然语言语句，该处理模块同时实现了收集模块 201 和自然语言处理模块 203 的功能，类似的变形仍在本发明的权利要求保护范围之内。

**[0063]** 图 3 所示的是信息分析流程图。需要的信息在步骤 301 从信息源 103（详见图 1）中被收集。信息源 103 可以包括但不限于服务器、通信终端。进一步地，服务器可以是 web 服务器、文件服务器、数据库服务器、FTP 服务器、应用程序服务器、代理服务器器等，或者上述服务器的任意组合。通信终端可以是手机、个人电脑、可穿戴设备、平板电脑、智能电视等，或者上述通信终端的任意组合。进一步地，在步骤 301，用户通过各种通信终端输入的自然语言语句可以被接收。上述需要的信息可以包括但不限于各种新闻、公告、评论、研报、博客、消息、报告、通知、论文、期刊中的一种或多种。上述需要的信息可以是关于各个行业的信息，包括但不限于体育、娱乐、经济、政治、军事、文化、艺术、科学、工程中的一种或多种。上述需要的信息的形式可以包括但不限于文字、图片、音频、视频等中的一种或多种。例如，新闻可以是某视频网站播放的视频新闻《世界银行下调今年全球经济增长预期至 2.8%》、某新闻网站报道的网页新闻《汇丰 5 月份中国服务业 PMI 升至 53.5》、某证券交易所发布的上市公司公告《A 股份有限公司关于签署日常经营重大合同的公告》、



某体育赛事直播平台发布的足球赛事预告《本周六切尔西俱乐部将在主场斯坦福桥球场迎战同城死敌阿森纳》等。步骤 301 可以由收集模块 201 完成。

**[0064]** 步骤 301 中收集的信息在步骤 302 被处理。步骤 302 可以由处理模块 202 完成。在本发明的一些实施例中，步骤 301 收集的信息可以为文字信息。该文字信息可以直接或间接来源于文本、音频、视频或上述来源的任意组合。进一步地，当文字信息来源于音频时，系统可通过语音识别或字幕提取将音频转化为文本。当文字信息来源于视频时，系统可通过语音识别或字幕文件提取将视频转化为文本。文字信息可以为汉语、英语、德语、西班牙语、阿拉伯语、法语、日语、韩语、俄语、葡萄牙语等，或上述语言的任意组合。进一步地，文字信息可以是字母、数字、字符、词语、短语、语句、段落、篇章等，或其中的一种或多种，或由任意数量的标识符组成的集合，该标识符集合可以包含一种或多种语义。步骤 302 执行的信息处理可以包括但不限于格式转换、分词处理、实体识别、数字及单位归一化处理、文本分类、事件属性抽取、细化事件识别等中的一种或多种。格式转换可以将各种格式的文字信息转换为统一的文本格式。文字信息的格式可以包括但不限于：pdf、doc、epub、mobi、caj、kdh、nh 等，或上述格式的一种或多种。统一的文本格式可以包括但不限于 txt、ASCII、MIME 等一种或多种组合。

**[0065]** 分词处理可以将文字信息中词语按照词语类型提取出来，词语类型可以包括但不限于名词、动词、形容词、副词、助词、拟声词、数字、专有符号等，或其中的一种或多种。可选择地，文字信息也可以应用一定的分词算法被处理。分词算法可以包括但不限于基于字符串匹配的分词方法（即机械性分词法）、基于理解的分词方法、基于统计的分词方法等，或上述分词方法的一种或多种。分词处理完成之后，对文本信息进行实体识别。实体可以包括但不限于产品、机构名、人名、地名、时间、日期、货币、数字、百分比等中的一种或多种。实体识别方法可以包括但不限于隐马尔科夫模型、最大熵模型、支持

向量机、基于规则的识别方法和基于统计的识别方法等，或其中的一种或多种。具体地，系统可以总结提炼过往信息中的要素，定义各种事件类别。例如：外交类、金融类、体育类、政治类、科学类、教育类等，或其中的一种或多种。上述类别也可包含若干级别的子类，如金融类可以包括股票、基金、期货等子类。上述类别可以包含实体识别完成之后，系统将对经过实体识别的文本信息中的数字和单位进行归一化处理。例如将“项目总投资三万元”转换为“项目总投资 30000 元”、将“梅西在巴萨主场对阵皇马的比赛中完成了帽子戏法”转换成“梅西在巴萨主场对阵皇马的比赛中打进 3 球”等。

5 [0066] 在完成数字和单位归一化处理之后，系统将对文本信息进行分类，以获取文本信息的大类别（如金融类）。在本发明的一个实施例中，系统可访问系统数据库 205（详见图 2），并将文本中出现的存储与数据库 205 中的类别关键词的数量或预设权值等属性特征通过一定的计算方法进行计算，并通过计算值进行分类。所述类别关键词可以通过特定的方法被提取出来，提取方法可以包括但不限于基于统计学的卡方统计、同义词规则、布尔关联规则、位置规则、信息增益、互信息、几率比、交叉熵、类间信息差等方法中的一种或多种的结合。在本发明的另一个实施例中，所述系统可采用基于机器学习的文本分类方法，包括但不限于决策树、Rocchio、朴素贝叶斯、神经网络、支持向量机、线性最小平方拟合、最邻近算法 kNN、遗传算法、最大熵等方法，或其中的一种或多种。即通过对标注好类别标签的文本的训练学习，得到分类器，从而对新的文本对象进行情感分类。文本信息经过分类可以被分到一个或多个类别中。类别可以是系统预定义的类别，并且可以包含若干级别的子类。如在金融领域，文本信息可以被分为包含但不限于公告类、新闻类、研报类、博客类、论坛类、微博类、互动投资类。公告类可以包含但不限于合同类、年报类、情感类等子类。可选择地，公告类又可以包括但不限于定期报告和权益分配公告、交易公告、募集资金公告、重大事项、政策优惠公告、高管人员变动公告、收购回购公告等。系统可以根据新闻来源将新闻分为

可靠来源和非可靠来源，如官方信息源央视财经频道可认为是可靠信息来源，新闻种类可以包括但不限于财经类、时政类、科教类、政法类、社会类、体育类、军事类、娱乐类等一种或多种组合。需要注意的是，上述对分类的描述只是为了便于理解发明，不应被视为是本发明唯一的实施例。分类步骤在本发明中不是必须的，对于一些文本信息，系统可以直接判定其类别，因而可以略过分类步骤。例如：对于某信息的标题显示为《A 公司关于签署日常经营重大经营合同的公告》，系统可直接判定其为公告类。

**[0067]** 完成文本分类之后，文本信息中的属性可以被抽取。属性为关于实体性质和关系的描述。例如，图 19 所示是《中信证券股份有限公司 2014 年年度报告》公告的封面和节选图，对于这一公告，抽取的实体可以为“中信证券”，从图 19 的表格中可以抽取出的属性包括营业收入、净利润同期增减、资产总额、负债总额、股东权益总额等，或其中的一种或多种。完成属性抽取之后，系统可以按照一定的规则方法，将实体与属性结合，生成细化事件。例如对于前述的实体“中信证券”，可以将其与其属性之一的“净利润同期增减”相结合可生成“中信证券 2014 年年报净利润增长 116.2%”这一细化事件。得到细化事件之后，系统可以根据得到的细化事件生成自然语言语句（步骤 303），比如“中信证券年报净利润增长大于 100%”，也可以是针对个股事件、行业事件和全市场事件的自然语言语句，比如“中信证券年报净利润同比增长率大于 100%，券商行业年报净利润同比增长率大于 100%，年报净利润同比增长率大于 100%。”。步骤 303 可以由自然语言处理模块 203 完成。生成的自然语言语句可以被输入到分析系统（例如，回测模块 204）中，以对识别出的事件进行分析（步骤 304）。在本发明的一个实施例中，步骤 304 可以由回测模块 204 完成。上述分析可以包括但不限于对事件进行回测。回测是指将事件与相关历史事件及数据按照一定的规则组合，生成回测报告，以供用户在投资时参考。

**[0068]** 需要注意的是，上述对信息分析系统流程的描述只是为了便

于理解发明，不应被视为是本发明唯一可行的实施例。系统也可以将收集到的信息直接转换为自然语言语句（步骤 303），然后对上述自然语言语句进行分析（步骤 304）。可选择地，系统也可以直接对收集到的信息进行分析（步骤 304）。

5 [0069] 图 4 所示的是收集模块 201 的结构示意图。收集模块 201 可以包括但不限于一个采集单元 401、一个处理单元 402、以及一个存储单元 403。采集单元 401 可以从信息源 103（详见图 2），或系统中的其他模块（例如，处理模块 202、自然语言处理模块 203、回测模块 204、系统数据库 205）采集需要的信息。上述需要的信息可以  
10 包括但不限于各种新闻、公告、评论、研报、博客、消息、报告、通知、论文、期刊等中的一种或多种。上述需要的信息可以是关于各个行业的信息，包括但不限于体育、娱乐、经济、政治、军事、文化、艺术、科学、工程等中的一种或多种。上述需要的信息的形式可以包  
15 括但不限于文字、图片、音频、视频等中的一种或多种例如，新闻可以是某视频网站播放的视频新闻《世界银行下调今年全球经济增长预期至 2.8%》、某新闻网站报道的网页新闻《汇丰 5 月份中国服务业 PMI 升至 53.5》、某证券交易所发布的上市公司公告《A 股份有限公司关于签署日常经营重大合同的公告》、某体育赛事直播平台发布的足球  
20 赛事预告《本周六切尔西俱乐部将在主场斯坦福桥球场迎战同城死敌阿森纳》等。可选择地，采集单元 401 也可以直接接收用户输入的信息，该信息可以包括但不限于自然语言语句、程序语言等。

[0070] 处理单元 402 可以对采集到的信息进行处理。处理可以包括但不限于将采集到的信息存入存储单元 403、将采集到的信息存入系统数据库 205、从存储单元 403 中调取信息并将信息发送给其他模块  
25 （例如，处理模块 202、自然语言处理模块 203、回测模块 204、系统数据库 205）、从系统数据库 205 中调取信息并将信息发送给其他模块（例如，处理模块 202、自然语言处理模块 203、回测模块 204）。可选择地，处理单元 402 也可将采集到的信息直接发送给其他模块，如处理模块 202、自然语言处理模块 203、回测模块 204、系统数据库

205。存储单元 403 可以存储收集模块 201 收集到的信息。存储单元 403 可以存储处理单元 402 处理过的信息。

[0071] 以上对收集模块的描述仅仅是具体的示例，不应被视为是唯一可行的实施方案。显然，对于本领域的专业人员来说，在了解所需的信息的基本原理后，可能在不背离这一原理的情况下，对所需要的信息的内容进行各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。

[0072] 图 5 所示的是处理模块 202 的结构示意图。处理模块 202 可以包含但不限于一个格式转换模块 501、一个文本预处理模块 502、一个文本分类模块 503、一个属性抽取模块 504、以及一个事件识别模块 505。上述各个模块可以是独立的，也可以是部分模块合并为一个模块。处理模块 202 中，格式转换模块 501 可以将收集模块 201 收集的信息进行格式转换。格式转换可以是系统自动完成的，也可以是人工完成的。格式转换可以是实时进行的，也可以是以固定的时间间隔进行的。可以转换的信息的文件格式包括但不限于 pdf、doc、docx、epub、mobi、caj、kdh、nh、bmp、jpg、tiff、gif、pcx、tga、exif、fpx、svg、psd、cdr、pcd、dxf、ufo、eps、ai、raw、mpeg、avi、mov、asf、wmv、navi、3gp、RA、RAM、mkv、flv、rmvb、WebM 等一种或多种组合。例如，收集的信息是图片 jpg 格式，如果该图片中包含文字信息，则可以通过 OCR (Optical Character Recognition) 识别将图片转换成文本格式，如 txt 格式。

[0073] 文本预处理模块 502 可以对格式转换后的文本进行预处理，预处理可以包括但不限于文本分词、实体识别、归一化处理等中的一种或多种。文本分类模块 503 可以对预处理后的文本进行分类，预处理后的文本可以被分为包括但不限于公告类、新闻类、研报类、博客类、论坛类、微博类、互动投资类。公告类可以包括但不限于合同类、年报类、等子类。可选择地，公告类又可以包括但不限于重组公告、股权激励公告、重大合同公告、政策优惠公告、高管变动公告、收购回购公告等。系统可以根据新闻来源将新闻分为可靠来源和非可靠来

源，如官方信息源央视财经频道可认为是可靠信息来源，新闻种类可以包括但不限于财经类、时政类、科教类、政法类、社会类、体育类、军事类、娱乐类等一种或多种组合。属性抽取模块 504 可以自动匹配抽取事件相关的属性，抽取规则可以是系统配置的，也可以是人工配置的。事件识别模块 505 可以根据属性抽取模块 504 的结果、文本预处理模块 502 的结果、系统数据库 205 以及一定的规则得出最终的细化事件。

**[0074]** 图 6 所示的是格式转换模块 501 的结构示意图。格式转换模块 501 可以包括但不限于一个控制单元 601、一个文本处理单元 602、一个图片处理单元 603、一个音频处理单元 604、以及一个视频处理单元 605。控制单元 601 可以根据收集模块 201 收集的信息选择相应的处理单元；文本处理单元 602 可以对收集模块 201 收集的文本格式的信息进行处理。图片处理单元 603 可以对收集模块 201 收集的图片格式的信息进行处理。音频处理单元 604 可以对收集模块 201 收集的音频格式的信息进行处理。视频处理单元 605 可以对收集模块 201 收集的视频格式的信息进行处理。本说明书所述实施例中上述单元可以是各自独立分布的，但在某些实施例中，上述部分单元可以合并为一个单元，如音频处理单元 604 可以和视频处理单元 605 合并为一个音视频处理单元，实现两者的功能。

**[0075]** 控制单元 601 可以对收集模块 201 收集的信息进行类型判断，并根据类型选择相应的处理单元。例如，控制单元 601 对收集模块 201 收集的信息进行判断后，判断是文本格式的信息，则为其选择文本处理单元 602 进行下一步骤的处理。

**[0076]** 文本处理单元 602 可以对收集模块 201 收集的文本格式的信息进行处理，转换为统一格式的文本数据。具体来说，收集模块 201 收集的信息中的文本格式可以包括但不限于超文本标识语言格式（Hypertext Markup Language, html）、可扩展超文本标识语言格式（Extensible Hypertext Markup Language, xhtml）、可扩展标识语言格式（Extensible Markup Language, xml）、pdf 格式（Portable Document

Format)、doc 及 docx 格式 (Microsoft 公司的专属格式) 等中的一种或多种, 文本处理单元 602 可以将上述格式转换为统一的文本格式, 统一的文本格式可以包括但不限于 txt 格式。例如, 收集模块 201 收集的信息是《中信证券股份有限公司 2014 年年度报告》, 如图 19 所示, 该公告的格式是 pdf 格式, 则文本处理单元 602 可以将该公告由 pdf 格式转换为 txt 格式。

[0077] 图片处理单元 603 可以对收集模块 201 收集的图片格式的信息进行处理, 转换为统一的文本格式。具体来说, 收集模块 201 收集的图片格式的信息可以是图书、报纸、杂志、信件等, 该类型的图片包含有文本信息在内, 图片处理单元 603 可以利用 OCR (Optical Character Recognition) 技术将图片信息转换为统一的文本格式。

[0078] 音频处理单元 604 可以对收集模块 201 收集的信息中的音频格式的数据进行处理, 转换为统一的文本格式。具体来说, 收集模块 201 收集的信息中音频格式可以包括但不限于 CD、WAVE、AIFF、AU、MPEG、MP3、MIDI、WMA、RealAudio、VQF、OggVorbis、AAC、APE 等中的一种或多种组合, 音频处理单元 604 可以利用语音识别技术将之转换为文本格式。语音识别技术可以包括但不限于基于声道模型和语音知识的方法、模式匹配的方法以及利用人工神经网络等方法中的一种或多种, 或是上述方法的任意组合。

[0079] 视频处理单元 605 可以对收集模块 201 收集的信息中的视频格式的数据进行处理, 转换为统一的文本格式。具体来说, 收集模块 201 收集的信息中的视频格式可以包括但不限于 Flash Video、AVI、WMV、MPEG、Mastroska、Real Video、QuickTime File Format、Ogg、MOD 等中的一种或多种组合, 视频处理单元 605 可以对视频中的字幕部分进行文本导出, 字幕包括视频内置字幕和外挂字幕, 并且转换为统一的文本格式。视频处理单元 605 也可以提取视频中的音频部分, 并进行语音识别将之转换为统一的文本格式。在具体的实施例中, 如果视频没有搭载字幕部分, 则视频处理单元 605 可以提取视频中的音频部分进行语音识别, 并转换为统一的文本格式, 如果视频搭载有字

幕，则将字幕导出并转换为统一的文本格式，也可以选择提取音频部分进行语音识别并转换为统一的文本格式。

**[0080]** 上述格式转换模块 501 包含的四种处理单元即文本处理单元 602、图片处理单元 603、音频处理单元 604 和视频处理单元 605，在一些实施例中可以不是全部包含的，可以只包含其中的一个单元，也可以是其中的某些单元。在一些实施例中，上述四种处理单元是全部包含的，各处理单元间的执行顺序可以是依次进行的，也可以是同时进行的，也可是任何合适的顺序。格式转换模块 501 对收集模块 201 收集的信息进行格式转换后转换为统一的文本格式，文本预处理模块 502 对该文本格式的信息进行后续处理。

**[0081]** 以上的描述仅仅是本发明的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明内容和原理后，都可能在不背离本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内。

**[0082]** 图 7 所示的是文本预处理模块 502 的结构示意图。文本预处理模块 502 可以包括但不限于一个语言识别单元 701、一个文本分词单元 702、一个实体识别单元 703、以及一个归一化单元 704。语言识别单元 701 可以对经过格式转换模块 501 处理后的文本信息进行语言识别。文本分词单元 702 可以对文本进行分词处理。实体识别单元 703 可以对文本中的实体进行识别。归一化单元 704 可以对文本中包含数字信息的内容及其对应的单位进行统一的归一化处理，形成标准的数字数据形式。上述单元可以是各自独立的，也可以是部分单元合并为一个单元。例如，语言识别单元 701 可以和文本分词单元 702 合并为一个单元。

**[0083]** 语言识别单元 701 可以对格式转换模块 501 处理后的文本进行语言识别。收集模块 201 收集的信息所用的语言可以包括但不限于汉语、英语、法语、俄语、西班牙语、阿拉伯语、日语、德语等中的一种或多种，语言识别单元 701 可以识别收集模块 201 收集的信息所



用的语言。

**[0084]** 文本分词单元 702 可以利用一定的分词算法对语言识别单元 701 识别后的文本进行分词处理。收集模块 201 收集的信息所用的语言中，包括以词为单位的语言，如英语、法语、俄语等，这些语言中词与词之间有天然的分隔；还包括以字为单位的语言，词是由字组成的，而词与词之间没有天然分隔，如汉语。因此在进行中文文本的词频统计前首先需要对中国文本进行分词处理，而英文文本则不需要。分词算法可以包括但不限于基于字符串匹配的分词方法（即机械性分词法）、基于理解的分词方法、基于统计的分词方法等，或是上述几种分词方法的任意组合。本发明的一个实施例为基于统计的分词方法和基于词典的分词方法相结合的方法。文本分词单元 702 可以向系统数据库 205 发送访问词典数据库的请求，系统数据库 205 收到请求时，可以将请求的词典发送给文本分词单元 702。词典可以是针对特定领域的词典，例如，可以是针对公告的词典或针对新闻的词典。具体地，该词典可以是针对重组公告的词典、针对激励公告的词典、针对重大合同公告的词典、针对政策优惠公告的词典、针对高管变动公告的词典、针对收购回购公告的词典等。文本分词单元 702 可以结合统计得到的分词结果和词典匹配得到的分词结果得到最终的文本分词结果。

**[0085]** 实体识别单元 703 可以通过实体识别方法对经过分词处理后的文本进行实体识别，可以将识别后的实体集合存入系统数据库 205 中的实体数据库。实体可以包括但不限于产品、机构名、人名、地名、时间、日期、货币、数字、百分比等中的一种或多种，具体举例来说，《中信证券股份有限公司 2014 年年度报告》这条标题信息中可识别的实体有“中信证券”、“股份有限公司”、“2014 年”、“年度报告”。实体识别方法可以包括但不限于隐马尔科夫模型、最大熵模型、支持向量机、基于布尔关联规则、基于同义词配置规则、基于位置规则的识别方法和基于统计的识别方法等，或是上述几种识别方法的任意组合。

**[0086]** 归一化单元 704 可以对文本中的数字及其单位进行归一化处

理，使其具有一致的单位。具体举例来说，归一化单元 704 可将文本中出现的“上涨概率百分之五”转换为“上涨概率 5%”，将“项目总投资三万元”转换为“项目总投资 30000 元”等。

**[0087]** 上述文本预处理模块 502 包含的四种处理单元执行顺序可以依次是语言识别单元 701、文本分词单元 702、实体识别单元 703 和归一化单元 704。文本预处理模块 502 中的处理单元的执行顺序也可以首先是语言识别单元 701 执行，依据语言识别单元 701 的识别结果来决定文本分词单元 702 是否执行。当识别结果是中文文本时，则执行文本分词单元 702。当识别结果是其他具有固定的分隔符的语言时，如英语、韩语、俄语等，则文本分词单元 702 可不执行。后续的实体识别单元 703 和归一化单元 704 的执行顺序可以是顺序的，可以是逆序的，也可以是同时进行的。文本预处理模块 502 可以对经格式转换模块 501 处理后的文本信息进行预处理，文本分类模块 503 可以对预处理后的文本进行后续处理。以上的描述仅仅是本发明的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明内容和原理后，都可能在不背离本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内。

**[0088]** 图 8 所示的是文本分类模块 503 的结构示意图。文本分类模块 503 可以包括但不限于一个或多个关键词提取单元 801、一个或多个分类单元 802。关键词提取单元 801 可以对文本预处理模块 502 处理后的文本进行关键词提取。分类单元 802 可以对提取出的关键词按照预先定义的规则对文本进行分类。上述单元可以是各自独立的，也可以是部分单元合并为一个单元。

**[0089]** 关键词提取单元 801 可以对文本预处理模块 502 处理后的文本进行分析并提取关键词。关键词的提取方法可以包括但不限于基于统计学的卡方统计、同义词规则、布尔关联规则、位置规则、信息增益、互信息、几率比、交叉熵、类间信息差等方法，或者上述方法的任意组合。具体来说，针对《中信证券股份有限公司 2014 年年度报

告》这一公告，关键词提取单元 801 首先进行关键词提取，提取的关键词可以包括但不限于“中信证券”、“2014 年”、“年度报告”、“净利润”、“同期增减”等等。

5 [0090] 分类单元 802 可以利用关键词提取单元 801 提出的关键词按照一定的分类方法对文本进行分类并贴上类别标签。分类方法可以包括但不限于决策树、Rocchio、朴素贝叶斯、神经网络、隐马尔科夫模型、支持向量机、线性最小平方拟合、最邻近算法 kNN、遗传算法、最大熵等方法，或者上述方法的任意组合。分类单元 802 可以向系统数据库 205 发送关键词数据库访问请求。系统数据库 205 收到请求后，  
10 将请求的关键词发送给分类单元 802。分类单元 802 可以依照一定的算法对关键词提取单元 801 提取的关键词和系统数据库 205 发送的关键词进行匹配，根据匹配结果来对文本分类，并给文本贴上相应的类别标签。具体来说，针对前述的题为《中信证券股份有限公司 2014 年年度报告》这一全文公告，分类单元将之归为公告大类，年度报告  
15 子类，并贴上标签。某个文本根据匹配结果，可同时归属于不同的类别，这时只要给该文本贴上两个标签即可，一个文本可以同时拥有两个以上的标签。

[0091] 文本分类模块 503 在某些实施例中是可选的。例如，如果收集模块 201 收集的信息中信息要素已经很确定，就可以略过文本分类  
20 步骤。具体地，如果收集的信息是一则新闻简讯，该新闻简讯的内容是“第 18 届上海国际电影节于 2015 年 6 月 21 日晚闭幕，主竞赛单元金爵奖揭晓，中国影片《烈日灼心》成最大赢家，邓超、郭涛、段奕宏三人同获影帝，曹保平获最佳导演”。这则简讯的要素如时间、人物、事件等就清晰明了，就可略过文本分类模块 503 直接传送至下一模块属性抽取模块 504 进行后续处理。上述文本分类模块 503 包含  
25 的关键词提取单元 801 和分类单元 802 的执行顺序可以是顺序的，即关键词提取单元 801 先执行，分类单元 802 后执行。以上的描述仅仅是本发明的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明内容和原理后，都可能在不背离

本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内。

**[0092]** 图 9 所示的是属性抽取模块 504 的结构示意图。属性抽取模块 504 可以包括但不限于一个或多个关键词提取单元 901、一个或多个属性抽取模板 902、一个或多个属性抽取单元 903。关键词提取单元 901 可以对文本进行关键词提取；属性抽取模板 902 可以存储能够抽取事件属性的抽取规则；属性抽取单元 903 可以完成对事件属性的抽取工作。上述单元可以是各自独立的，也可以是部分单元合并为一个单元。

**[0093]** 关键词提取单元 901 可以对文本预处理模块 502 处理后的文本进行分析并提取关键词，提取方法可以包括但不限于基于统计学的卡方统计、同义词规则、布尔关联规则、位置规则、信息增益、互信息、几率比、交叉熵、类间信息差等方法，或者上述方法的任意组合。关键词提取单元 901 在某些实施例中是可选的。由于文本分类模块 503 是可选的，当需要略过文本分类模块 503 直接由属性抽取模块 504 对信息进行处理时，关键词提取单元 901 可以被执行，对预处理后的文本进行关键词的提取。如果文本分类模块 503 已经执行过，则可略过关键词提取单元 901。

**[0094]** 属性抽取模板 902 可以存储能够抽取事件属性的抽取规则。事件是由实体及实体的属性组成的。文本预处理模块 502 中的实体识别单元 703 已经对文本进行过实体识别，并形成了实体集合，属性抽取模板 902 存储的是针对不同实体的不同的属性抽取规则。抽取规则是预先配置的。配置方法可以是人工配置，根据预先设定的文本的种类，为每个种类的文本设定不同的实体属性抽取规则。配置方法也可以是机器学习方法。例如，可以首先选定一批训练文本。该训练文本是通过人工标注的类别清晰的文本。通过对训练文本的训练学习，得到属性抽取器（图中未展示）。该属性抽取器可根据不同的文本类别抽取所需的属性，然后利用该属性抽取器对新的文本进行属性抽取。属性抽取模板 902 在某些实施例中是可选的。当文本分类模块 503 没

有执行，文本没有分类就没有对应的属性抽取模板。

[0095] 属性抽取单元 903 可以完成从文本中抽取属性的工作。当文本分类模块 503 执行后，属性抽取单元 903 就可以根据文本分类的结果及文本所贴标签来选择相应的属性抽取模板。当文本贴有两个及以上  
5 的标签时，可以同时选取对应数量的属性抽取模板，然后对文本进行属性抽取，将所得结果进行聚类。具体举例来说，对于《中信证券股份有限公司 2014 年年度报告》这一信息，文本分类模块 503 赋予它的标签是公告大类年度报告子类。依据该标签，属性抽取单元 903 选取相应的属性抽取模板，根据选取的模板抽取事件属性。事件属性  
10 可以包括但不限于营业收入金额及同比涨幅大小、净利润金额、总资产金额及涨幅大小等等方面的信息。举例来说，对上述公告可抽取其中的“归属于母公司股东的净利润/本期比上年同期增长(%) /116.20%”这条属性信息。以上的描述仅仅是本发明的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明  
15 内容和原理后，都可能在不背离本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内。

[0096] 回到图 5，事件识别模块 505 可以完成对事件的识别工作，文本预处理模块 502 中的实体识别单元 703 已经对文本进行过实体识别，并形成了实体集合。属性抽取模块 504 从文本中抽取出了所需的  
20 属性集合。事件识别模块 505 可以依据实体识别结果和属性抽取结果，依据一定的事件识别模板，将事件识别出来，并生成细化事件。具体举例来说，对于《中信证券股份有限公司 2014 年年度报告》，前述文本预处理模块 302 识别的实体包括“中信证券”，“股份有限公司”，  
25 “2014 年”，“年度报告”，“净利润”，“同期增减”等等，属性抽取模块 304 抽取的一条事件属性是：“归属于母公司股东的净利润/本期比上年同期增减(%) /116.20%”，事件识别模块 505 可以根据实体识别结果和属性抽取结果得到最终的细化事件：“中信证券 2014 年年度报告公告归属于母公司股东的净利润同比增长率等于 116.2%”。

**[0097]** 在某些实施例中，当收集模块 201 收集到用户输入的信息，对该信息进行回测时，可能会存在复杂逻辑，事件识别模块 505 仅仅根据实体信息和属性信息不能细化地识别出事件。这时需要根据系统数据库 205 中的事件库(事件属性数据库 1405 或事件识别数据库 1411)和一定的规则方法来进行事件的识别。具体举例来说，当用户输入的信息是“中标金额占营业收入 50%以上合同的公司”，那么中标公告中很可能不会包含该公司营业收入的数据。这时可以根据数据库的数据(如历史数据)和一定的规则方法来计算得到最终的细化事件类别。以上的描述仅仅是本发明的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明内容和原理后，都可能在不背离本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内

**[0098]** 图 10 所示的是处理模块的流程图。系统对收集模块 201 传输的信息进行格式转换，将之转换为统一的文本格式(步骤 1001)。格式转换可以包括但不限于对文本、图片、音频、视频等中的一种或多种进行格式转换。步骤 1001 可以由格式转换模块 501 实现。系统对文本格式的信息进行预处理(步骤 1002)。预处理可以包括但不限于语言识别、文本分词、实体识别、归一化等中的一种或多种。步骤 1002 可以由文本预处理模块 502 实现。系统对文本进行分类(步骤 1003)。分类步骤可以包括但不限于进行关键词提取和进行分类。步骤 1003 可以由文本分类模块 503 实现。系统对文本进行属性抽取(步骤 1004)。步骤 1004 可以由属性抽取模块 504 实现。系统对事件进行识别(步骤 1005)。步骤 1005 可以由事件识别模块 505 实现。可选地，系统还可以不经过步骤 1001 直接进行步骤 1002，不经过步骤 1003 直接进行步骤 1004。

**[0099]** 图 11 所示的是自然语言处理模块 203 的结构示意图。自然语言处理模块 203 可以包括但不限于一个收集单元 1101，以及一个自然语言生成单元 1102。收集单元 1101 可以通过访问系统中的其他模

块收集需要的信息（例如，收集模块 201、处理模块 202、回测模块 204、系统数据库 205）。自然语言生成单元 1102 可以将收集单元 1101 收集到的信息转化为自然语言语句。在本发明的一些实施例中，收集单元 1101 可以接收处理模块 202 输出的细化事件。同时，收集单元 5 1101 还可以从收集模块 201 接收用户输入信息。自然语言生成单元 1102 可以接收细化事件，并根据用户输入信息，对细化事件进行处理。在本发明的一种实施例中，例如，在股票领域，针对公告，用户可以选择生成针对个股的自然语言语句，或者选择针对行业的自然语言语句，还可以选择针对全市场的自然语言语句，或者其中的一种或多种。

10 例如在新闻领域，针对 IPO 新股，用户可以选择生成针对大盘的自然语言语句。需要注意的是，上述的自然语言语句也可以在用户不介入的情况下自动生成。例如，针对个股，系统可以自动生成针对个股的自然语言语句、针对行业的自然语言语句、针对全市场的自然语言语句，或者其中的一种或多种的结合。在新闻领域，针对 IPO 新股，系

15 统可以自动生成针对大盘的自然语言语句。进一步地，在新闻领域，自然语言生成单元 1102 可以生成针对商品价格的自然语言语句、针对天气状况的自然语言语句、针对人口统计的自然语言语句。上述自然语言语句中相关的数值可以缺省（在没有数值的情况下只会陈述改事件是否发生）。自然语言生成单元 1102 生成的自然语言语句可以被

20 输入到回测模块 204 中，以对该自然语言语句进行回测。在本发明的另一些实施例中，用户输入待回测的自然语言语句到收集模块 201，自然语言处理模块 203 从收集模块 201 中接收用户输入的自然语言语句。可选择地，用户可以直接将待回测的自然语言语句输入到自然语言处理模块 203 的查询框中，（图中未展示）。自然语言处理模块 203

25 可以对用户输入的自然语言语句进行预处理，得到标准节点序列（节点至少包括指标节点和条件节点），并根据指标节点和其他节点之间的关系，构造节点树。节点树可以用于表征指标条件组合。根据节点树可以生成数据查询指令。该数据查询指令可以被输入回测模块 204 中以进行回测分析。用户可以通过自然语言处理模块 203 调用存储在

系统数据库 205 中的历史数据，用户可以利用布尔操作符（AND、OR、NOT 等）将一定数量的自然语言语句组合到一起。

[0100] 需要注意的是，以上对自然语言处理模块 203 的描述仅仅是具体的示例，不应被视为是唯一可行的实施方案。显然，对于本领域的专业人员来说，在了解所需要的信息的基本原理后，可能在不背离这一原理的情况下，对所需要的信息的内容进行各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。例如，自然语言处理模块 203 也可以接收回测模块 204 输出的回测结果。

[0101] 图 12 所示的是回测模块的结构示意图。回测模块 204 可以包含一个标准问句单元 1201、一个其他问句单元 1202、一个优化单元 1203、和一个扩展单元 1204。标准问句单元 1201、其他问句单元 1202、优化单元 1203 和扩展单元 1204 可以是独立的。上述单元中的一些单元也可以是合并为一个单元工作。

[0102] 标准问句单元 1201 可以接收系统标准的自然语言语句事件。本发明的一些实施例中，标准问句单元 1201 可以接收自然语言处理模块 203 生成的自然语言语句，也可以接收来源于其他模块的自然语言语句。其他模块包括但不限于收集模块 201、处理模块 202、自然语言处理模块 203、和系统数据库 205 等一种或多种组合。其他问句单元 1202 可以接收非系统的自然语言语句。非系统的自然语言语句可以包括但不限于用户输入、专家定义、系统抽取的结果等一种或多种组合。标准问句单元 1201 和其他问句单元 1202 可以合并为一个问句单元，该问句单元可以接收系统和用户输入的自然语言语句事件等。

[0103] 优化单元 1203 可以根据接收的信息和回测算法优化出策略组合。优化方式可以是自动的，也可以是人工的。例如，在金融领域，上述的回测算法得到的基本回测数据包括但不限于持有期、单次收益平均值、单次收益最大值、单次收益最小值、预期年化收益率、交易次数、盈亏比、成功率、最大回撤率、周战胜率、夏普比率、最大连续无选股结果天数、平均每天选股数等一种或多种组合。优化单元 1203 进一步还可以根据回测结果给出最优策略，同时还会有报告评级



等。扩展单元 1204 可以配置提供订阅功能，也可以配置提供信息分享功能。订阅功能可以是扩展单元 1204 根据用户选择订阅包含特定关键词的信息，通过各种方式将经过该系统分析后的信息内容推送给用户。分享功能可以是用户通过各种方式把感兴趣的信息分享给朋友等。

**[0104]** 图 13 所示的是回测流程图。步骤 1301 接收信息。接收的信息来源可以包括但不限于收集模块 201、处理模块 202、自然语言处理模块 203、和系统数据库 205。接收的信息可以是自然语言语句，也可以是机器语句。步骤 1301 可以是回测模块 204 接收到自然语言语句。自然语言语句信息可以由用户直接输入的，也可以是其他模块生成的。步骤 1301 可以由标准问句单元 1201 和/或其他问句单元 1203 完成。

**[0105]** 步骤 1302 中，将接收到的自然语言语句与历史数据进行回测分析，历史数据可以储存在系统数据库 205，也可以储存在回测模块 204。自然语言语句与历史数据的回测分析可以通过一定的优化方法实现。优化方法可以包括但不限于系统定义、用户自定义选择、机器学习等其中的一种或多种组合。

**[0106]** 步骤 1303 中，优化分析的信息结果会与相应的文字模板匹配。文字模板可以是系统定义的，也可以是用户自定义的。与模板匹配的结果内容可以包括但不限于回测报告、报告评级、最优策略、走势预测等一种或多种组合。

**[0107]** 需要注意的是，上述对信息分析系统流程的描述只是为了便于理解发明，不应被视为是本发明唯一可行的实施例。显然，对于本领域的专业人员来说，在了解所需要的基本原理后，可能在不背离这一原理的情况下，对所需要的信息的内容进行各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。例如，步骤 1303 中，回测的历史信息可以与文字模板匹配，也可以是与语音、视频、图片等模板匹配。又如，回测过程可以由回测模块 204 完成，也可以由自然语言处理模块 203、处理模块 202、收集模块 201 等完成。

[0108] 回到图 12, 回测模块 204 进一步可以包括扩展单元 1204。扩展单元 1204 可以配置提供订阅功能, 也可以配置提供信息分享功能。扩展单元可以包括但不限于各种类型的应用程序接口 (API), 如面向对象的 API、库与框架的 API、API 与协议、API 与设备接口、web API, 或其中的一种或多种。订阅功能可以是扩展单元 1204 根据用户选择订阅包含特定关键词的信息, 通过各种方式将经过该系统分析后的信息内容推送给用户。分享功能可以是用户通过各种方式把感兴趣的信息分享给朋友。扩展单元 1204 的订阅功能可以包括但不限于为用户提供推送信息, 也可以推荐关注兴趣相似的用户, 还可以推荐信息的评论, 并且提供信息有无帮助的评分等。扩展单元 1204 推送的方式可以包含但不限于移动客户端软件、电子邮件、短信、RSS 门户网站、在线单用户聚合器、搜索引擎、浏览器、即时通讯软件、社交网络等。扩展单元 1204 的推送周期可以是系统设定的, 也可以是用户自定义的。推送周期可以是定期的也可以是不定期的。推送可以是实时的也可以是延时的。扩展单元 1204 推送的信息内容形式可以包括但不限于文字、语音、图片、动画、视频等中的一种或多种。扩展单元 1204 推送的信息内容可以包括但不限于用户已浏览的信息内容更新, 可以是用户关注的信息, 也可以是系统根据用户记录推荐的信息, 还可以是同类信息关注的热度情况等中的一种或多种。

[0109] 扩展单元 1204 的分享功能可以是用户使用的一种发布信息方式, 分享到指定的地方, 选择哪些人可以看到该信息等。信息分享的内容可以是单条信息也可以是多条信息, 可以是部分选取内容的信息也可以是页面整体内容的信息, 可以是信息内容分享也可以是信息评论分享, 可以是信息的关注度分享也可以是信息的帮助评分分享等。信息分享的方式可以包括但不限于短信、彩信、电子邮件、QQ、MSN、微信、微博、豆瓣、Twitter、Facebook、Instagram、人人、即时通讯软件工具等中的一种或多种。信息分享接收对象可以包括但不限于单个朋友、多个朋友、朋友圈、公共社交圈、论坛、其他用户等中的一个或多个。信息分享的内容格式可以包括但不限于文字、图片、语音、

动画、视频、网页链接等中的一种或多种。以上对信息分享模式其所实现功能的描述仅仅是具体的示例，不应被视为是唯一可行的实施方案。以上对扩展单元 1204 的描述仅仅是具体的示例，不应被视为是唯一可行的实施方案。显然，对于本领域的专业人员来说，在了解扩展单元 1204 的基本原理后，可能在不背离这一原理的情况下，对实施扩展单元的具体方式与步骤、以及扩展单元所能实现的功能进行形式和细节上的各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。

[0110] 图 14 所示的是系统数据库 205 的模块示意图。系统数据库 205 可以包括但不限于一个原始信息数据库、一个文本数据库、一个文本预处理数据库、一个实体数据库、一个事件属性数据库、一个关键词数据库、一个文本分类数据库、一个历史信息数据库、一个自然语言处理数据库、一个事件识别数据库、一个回测模块数据库、一个文字模板数据库、一个词典数据库等一种或多种组合。系统数据库 205 可以储存数据及模板，也可以处理数据。例如，历史信息数据库 1409 收集的历史信息可以在该数据库中进行分类储存。同样地，当信息处理过程中有任何的更新，也会在各数据库中进行信息实时更新，如关键词数据库 1406，同义词的更新实现在该数据库。显然，对于本领域的专业人员来说，在了解信息分析系统及方法的数据库原理后，可能在不背离这一原理的情况下，对实施上述方法和系统的应用领域形式和细节上的各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。例如，系统数据库 205 中的各类数据库可以是在收集模块 201、处理模块 202、自然语言处理模块 203、和回测模块 204 中分别实现其功能。系统数据库 205 中的各类数据库也可以是一种数据库实现两种及两种以上数据的库功能，如文本预处理数据库 1403 可以同时储存预处理数据、实体数据、事件属性和关键词等，同时实现实体数据库 1404、事件属性数据库 1405、和关键词数据库 1406 的功能。

[0111] 图 15 所示的是信息分析流程图。信息分析系统在步骤 1501 对需要的信息进行收集。步骤 1501 可以由收集模块 201 完成。上述

需要的信息可以包括但不限于各种新闻、公告、评论、研报、博客、消息、报告、通知、论文、期刊等中的一种或多种。上述需要的信息可以是关于各个行业的信息，包括但不限于体育、娱乐、经济、政治、军事、文化、艺术、科学、工程等中的一种或多种。上述需要的信息的形式可以包括但不限于文字、图片、音频、视频等中的一种或多种。在本发明的一些实施例中，系统在步骤 1501 收集的信息可以是文本信息。该文本信息包括但不限于以下格式：pdf、doc、epub、mobi、caj 等，或其中的一种或多种。

**[0112]** 系统在步骤 1502 可以对文本信息进行预处理。步骤 1502 可以由处理模块 202 完成。文本预处理可以包括但不限于格式转换、分词处理、实体识别、数字及单位归一化处理等其中的一种或多种组合。例如，系统在步骤 1501 收集到的信息为“中信证券股份有限公司 2014 年年度报告”公告，该公告为 PDF 文件，可以在上海证券交易所网站下载。系统通过格式转换，将该公告转换成 txt 格式的文本，以方便分词和后续的文本处理。同时对 PDF 格式内部的表格部分进行分析处理，保留部分格式化信息和上下文信息。格式转换完成后，系统将按照一定的方法对该公告进行分词处理。分词处理可以依据一个统计模型以及词典数据库 1407 进行。可选择地，分词处理也可通过应用一定的规则实现。规则可以包括但不限于，同义词配置、布尔关联规则、位置规则，或其中的一种或多种。完成分词处理之后，系统将对公告进行实体识别。实体识别包括但不限于产品、机构名、人名、地名、时间、日期、货币、数字、百分比等。具体地，系统可以总结提炼过往信息中的要素，定义各种事件类别。例如：外交类、金融类、体育类、政治类、科学类、教育类等，或其中的一种或多种。上述类别也可包含若干级别的子类，例如，金融类包含国债类、股票类、基金类等。完成识别后，系统将对该公告进行数字及单位归一化处理。例如，将“净利润增长百分之三”转换为“净利润增长 3%”。

**[0113]** 在完成本文预处理之后，系统将对该公告进行文本处理（步骤 1503）。文本处理可以由处理模块 202 完成。系统对经过预处理后

的公告进行关键词匹配。关键词匹配可以与同义词配置、布尔关联规则、位置规则等方法相结合。根据关键词匹配或其它方法处理的结果，系统可以对该公告进行类别判断（步骤 1504）。例如，图 19 所示的《中信证券股份有限公司 2014 年年度报告》，抽取出的关键词可以为“中信证券”“2014 年”“年度报告”“净利润”“同期增减”等，则该公告可以被判定为财务报告大类。步骤 1504 可以由处理模块 202 完成。

**[0114]** 在完成类别判断之后，系统可以依据一个规则生成细化事件（步骤 1505）。在本发明的一些实施例中，完成类别判断后，文本信息中的属性可以被抽取出来。属性为关于实体性质或关系的描述。例如，对于图 19 所示的《中信证券股份有限公司 2014 年年度报告》，抽取出的实体可以为“中信证券”，则从图 19 的表格中可以抽取出的属性可以是营业收入、净利润同期增减、资产总额、负债总额、股东权益总额等，或其中的一种或多种。完成属性抽取之后，系统可以按照一定的规则方法，将实体与属性结合，生成细化事件。步骤 1505 可以由处理模块 202 完成。生成细化事件的规则可以包括但不限于，同义词配置、布尔关联规则、位置规则，或其中的一种或多种。对于上述公告所抽取出的实体“中信证券”，可以将其与其属性之一的“净利润同期增减”相结合可生成“中信证券 2014 年年报净利润增长 116.2%”这一细化事件。

**[0115]** 生成细化事件后，系统可以针对该细化事件生成自然语言语句（步骤 1506）。步骤 1506 可以由自然语言处理模块 203 完成。对于上述公告，系统可以生成 3 句自然语言语句，“中信证券年报净利润同比增长大于 100%”，“券商行业年报净利润同比增长率大于 100%”，“年报净利润同比增长率大于 100%”，即分别对应个股、行业和整个股票市场等三个层次。系统可以针对以上生成的 3 句自然语言语句，分别进行个股事件回测、行业事件回测、以及全市场回测（步骤 1507）。步骤 1507 可以由回测模块 204 完成。完成回测之后，系统可以将回测结果匹配文字模板，生成回测报告（步骤 1508）。步骤 1508 可以由

回测模块 204 完成。对于上述公告，系统生成的一个示例回测报告可以是“近 1 年，A 股所有财务报告公告，次日收盘平均收益 0.48%，上涨概率 47.77%。其中，证券行业共公布 165 次同类公告，次日收盘平均收益 0.72%，上涨概率 46.67%，股价次日下跌概率偏大，获利

5 概率偏低。最优策略：上涨概率最高是持股 11 天后收盘卖出，平均收益 9.40%。”由于上涨概率分布在 50% 上下，因此判定“无足轻重”。

[0116] 需要注意的是，上述对信息分析系统流程图的描述只是为了便于理解发明，不应被视为唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明的内容和原理之后，都可能在不背离本发明

10 原理和结构的情况下，进行形式和细节上的各种修正和改变，但这些修正和改变仍在本发明的权利要求保护范围之内。例如，在步骤 1504，系统可以直接收集用户输入的信息，并将该信息转换为自然语言语句（步骤 1506）。同样地，系统在完成信息收集之后，可以直接进行文本处理（步骤 1503）。同样地，系统在完成文本处理之后，可以直接

15 依据处理后的文本生成细化事件，步骤 1504 不是必须的。

[0117] 图 16 所示的是信息分析系统交互工作流程图。信息分析系统在步骤 1601 接收自然语言语句。该自然语言语句可以是用户直接输入的，也可以是通过处理公告、新闻等文本得到的自然语言语句。具体地，用户可以通过信息分析系统提供的交互界面（详见图 18）

20 来输入自然语言语句。用户可以输入任何自然语言语句。例如，在金融领域，用户可以输入“000826，签订重大合同”；系统可以根据输入的自然语言语句检索代码 000826 对应的公司签订的重大合同。自然语言语句的数字和日期可以是任意格式的（详见图 18）。

[0118] 信息分析系统在步骤 1602 处理步骤 1601 中接收的自然语言

25 语句。处理可以包括但不限于，分词处理、实体识别、数字及单位归一化处理、文本分类、事件属性抽取、细化事件识别等一种或多种组合。信息分析系统在步骤 1603 对步骤 1602 中经过处理的自然语言语句进行回测，然后在步骤 1604 生成回测报告。关于回测报告的具体内容将在图 18 中作详细描述。

[0119] 图 17 是信息分析系统针对新闻或公告的一个交互界面示意图。参考图 2，该交互界面可以由回测模块 204 生成并用来展示回测结果（回测展示）。该交互界面可以在各种电子设备上显示。电子设备可以包括但不限于手机、个人电脑、平板电脑、PDA、智能手表、  
5 智能家电、智能交通工具等，或其中的一种或多种。在地址栏中，用户可以输入任意公告或新闻的统一资源定址器（URL）以对该公告或新闻及其分析结果进行阅读。可选择地，用户也可以输入任意公告或新闻的 IP（Internet Protocol）地址。

[0120] 在查询框 1701 中，用户可以输入公告或新闻的完整名称。

10 可选择地，用户也可以输入公告或新闻中的关键字，以在结果列表中选择指定的新闻或公告。在选定好指定公告或新闻之后，该公告或新闻的标题、正文内容、以及回测报告可以在交互界面上被显示出来。区域 1702 可以显示公告或新闻的正文内容的全部或部分，用户可以通过鼠标、按键、触摸屏、语音控制或者触摸板来查看正文内容。区  
15 域 1703 可以显示针对选定公告或新闻的回测报告。所述回测报告可以包括但不限于对历史数据的展示，例如，

[0121] “近 1 年，A 股所有签订重大合同公告，次日收盘平均收益 0.38%，上涨概率 47.62%。

20 [0122] 其中，XX 制药共公布 16 次同类公告，次日收盘平均收益 0.29%，上涨概率 68.75%，股价次日上涨概率极大，获利概率较高。”

[0123] 除了展示历史数据外，对公告或新闻的评级和建议策略也可以被展示在区域 1703 中。所述建议策略基于回测周期中最晚时间之后的最优历史表现，建议策略可以是，例如，上涨概率最高是持股 1 天后收盘卖出，平均收益 0.29%。报告评级可以是利好、利空或无足  
25 轻重等。区域 1704 可以显示最新公告或新闻以方便用户查看。显示方式可以是列表显示，即只显示最新公告或新闻的标题和时间。区域 1705 可以显示选定报告或新闻的一些相关信息，如选定报告或新闻的，公告（新闻）类型、发布时间、新闻或公告中涉及的证券名称、证券代码、公告（新闻）编号等。

[0124] 以上的描述仅仅是本发明分类展示模块的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明内容和原理后，都可能在不背离本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内。

[0125] 图 18 是信息分析系统针对用户输入的一个交互界面示意图。参考图 2，该交互界面可以由回测模块 204 生成并用来展示回测结果（回测展示）。该交互界面可以在各种电子设备上显示，电子设备可以包括但不限于，手机、个人电脑、平板电脑、PDA、智能手表、智能家电、智能交通工具等，或其中的一种或多种。在地址栏中，用户可以输入任意文本的统一资源定址器（URL）以对该文本及其分析结果进行阅读。可选择地，用户也可以输入任意文本的 IP（Internet Protocol）地址。

[0126] 在区域 1801 中，用户可以输入任何自然语言语句。如在金融领域，用户可以输入“12 月 10 日 20 日均线粘合；振幅小于 3%；两比从大到小排序”或者“股息率连续两年大于 3%；每股收益大于 2 元；市值小于 50 亿营业收入同比增长率从小到大”。输入的自然语言语句中的数字和日期可以是以任意格式表现的数字和日期。日期可以是前大前天、上上个星期、上个周末、上周 5、上个月、上个季度、大前年、近 N 天、近 N 周、近 N 个交易日等。数字可以是 3 分之 1、1/3、5 元、百分之 5、5% 等。数字也可以是一个范围，如 5 至 10 元、5-10% 等。输入的自然语言语句中还可以加入排序规则。排序规则可以是，如量比从大到小、量比从小到大、涨跌幅从小到大、涨跌幅从大到小、换手率从大到小、换手率从小到大、资金流向从大到小、流通盘从小到大、DDE 从大到小、市值从小到大、基本每股收益从大到小、销售毛利率从大到小、销售毛利率从大到小、净资产（同比增长率）从大到小、资产收益率 roe 从大到小、已获利息倍数从大到小、营业收入（同比增长率）从大到小等，或其中的一种或多种。

[0127] 在区域 1802，用户可以设置分析策略。例如，用户可以设置



时间范围、持仓股票、买入时机、持股周期、止盈条件、止损条件、交易费率等。具体地，用户可以设置买入时机为第二天开盘后买入，也可以设置止盈条件为“大于 25% 时，回撤 5% 止盈”。在设置好区域 1801 以及区域 1802 中的内容后，用户可以点击查询按钮，关于输入自然语言语句的回测报告将被生成。区域 1803 可以展示对生成回测报告的评级以及建议的策略。报告评级可以是对最大预期年化收益率及最大成功率的估计等。区域 1804 可以显示针对输入的自然语言语句的回测报告。回测报告可以包括但不限于回测数据分析、累计收益图、收益分布图、历史交易查询等内容。回测数据可以包括但不限于持有期、单次收益平均值、单次收益最小值、预期年化收益率、交易次数、盈亏比、成功率、最大回撤率、周战胜率、夏普比率、最大连续无选股结果天数、平均每天选股数等

**[0128]** 以上的描述仅仅是本发明分类展示模块的具体实施例，不应被视为是唯一的实施例。显然，对于本领域的专业人员来说，在了解本发明内容和原理后，都可能在不背离本发明原理、结构的情况下，进行形式和细节上的各种修正和改变，但是这些修正和改变仍在本发明的权利要求保护范围之内。

**[0129]** 以上对适用领域的描述仅仅是具体的示例，不应被视为是唯一可行的实施方案。显然，对于本领域的专业人员来说，在了解一种信息分析方法和系统的基本原理后，可能在不背离这一原理的情况下，对实施上述方法和系统的应用领域形式和细节上的各种修正和改变，但是这些修正和改变仍在以上描述的范围之内。凡是能整理成数据的配置系统都可以使用本发明描述的系统实现信息分析的功能，例如本发明可以作为浏览器插件，当用户浏览网站时，需要对当前网页的新闻或公告进行信息分析，可使用该插件回测该新闻或公告的历史信息，并给予预测；同样地，该系统还可以嵌入公司系统中对财务报表进行智能数据分析；另外，各种传感器采集数据，如温度传感器、湿度传感器、风力传感器可以读取环境数据，可以通过该系统来分析环境历史趋势并预测未来环境变化；医学方面，不同年龄段使用同一药物的

效果进行回测，如感冒得病的症状，根据历史数据分析得出多少天痊愈等。

## 权 利 要 求 书

1. 一种信息分析系统，包括：

一种计算机可读的存储媒介，所述存储媒介存储可执行模块，  
包括：

5 收集模块，所述收集模块能够收集信息；

处理模块，所述处理模块能够对收集的信息进行预处理，  
从预处理后的信息中提取事件；

自然语言处理模块，所述自然语言处理模块能够根据提取  
出的事件生成自然语言语句；

10 回测模块，所述回测模块能够根据生成的自然语言语句获  
取历史信息，并结合所述历史信息生成回测结果；

一个处理器，所述处理器能够执行所述计算机可读的存储媒介  
存储的可执行模块。

2. 根据权利要求 1，所述的信息分析系统进一步包括一个数据库，  
15 所述数据库能够储存所述的收集信息、预处理后的信息、提取的事件、  
自然语言语句、历史信息、回测结果。

3. 根据权利要求 2，所述的数据库包括原始信息数据库、文本数  
据库、文本预处理数据库、实体数据库、事件属性数据库、关键词数据  
库、文本分类数据库、历史信息数据库、自然语言处理数据库、事件识  
20 别数据库、回测模块数据库、文字模板数据库、词典数据库。

4. 根据权利要求 1，所述的处理模块进一步包括格式转换模块、  
文本处理模块、属性抽取模块、事件识别模块。

5. 根据权利要求 4，所述的处理模块进一步包括文本分类模块。

6. 根据权利要求 1，所述的处理模块采用的方法包括卡方统计、  
25 信息增益、互信息、几率比、交叉熵、类间信息差、关键词统计、决策  
树、Rocchio、朴素贝叶斯、神经网络、支持向量机、线性最小平方拟合、  
最邻近算法 kNN、遗传算法、情感分类、最大熵、Generalized Instance Set、  
同义词配置、布尔关联规则、位置规则、机器学习。

7. 根据权利要求 1，所述的自然语言处理模块可以从收集模块接

收信息。

8. 根据权利要求1, 所述的回测模块进一步包括回测信息判断, 所述回测信息判断根据回测结果的情况给出评价。

9. 根据权利要求1, 所述的回测结果可以被展示给用户。

5 10. 一种信息分析方法, 包括:

收集信息;

根据所述信息提取事件;

根据所述事件生成自然语言语句;

根据所述自然语言语句获取历史信息;

10 结合所述历史信息对所述自然语言语句进行回测分析。

11. 根据权利要求10, 所述的收集信息包括收集用户输入信息和非用户输入信息, 所述非用户输入信息来源包括通信终端和服务器。

12. 根据权利要求11, 所述的收集信息包括收集公告信息和新闻信息。

15 13. 根据权利要求10, 所述的提取事件进一步包括实体识别和属性抽取。

14. 根据权利要求13, 所述的实体识别进一步包括格式转换、文本分词、数字及单位归一化处理。

20 15. 根据权利要求13, 所述的属性抽取可以通过系统定义的模型实现。

16. 根据权利要求10, 所述的自然语言语句可以根据用户输入信息生成。

17. 根据权利要求10, 所述的自然语言语句可以进一步根据事件类别扩展。

25 18. 根据权利要求10, 所述的自然语言语句回测可以根据信息类别生成回测结果。

100

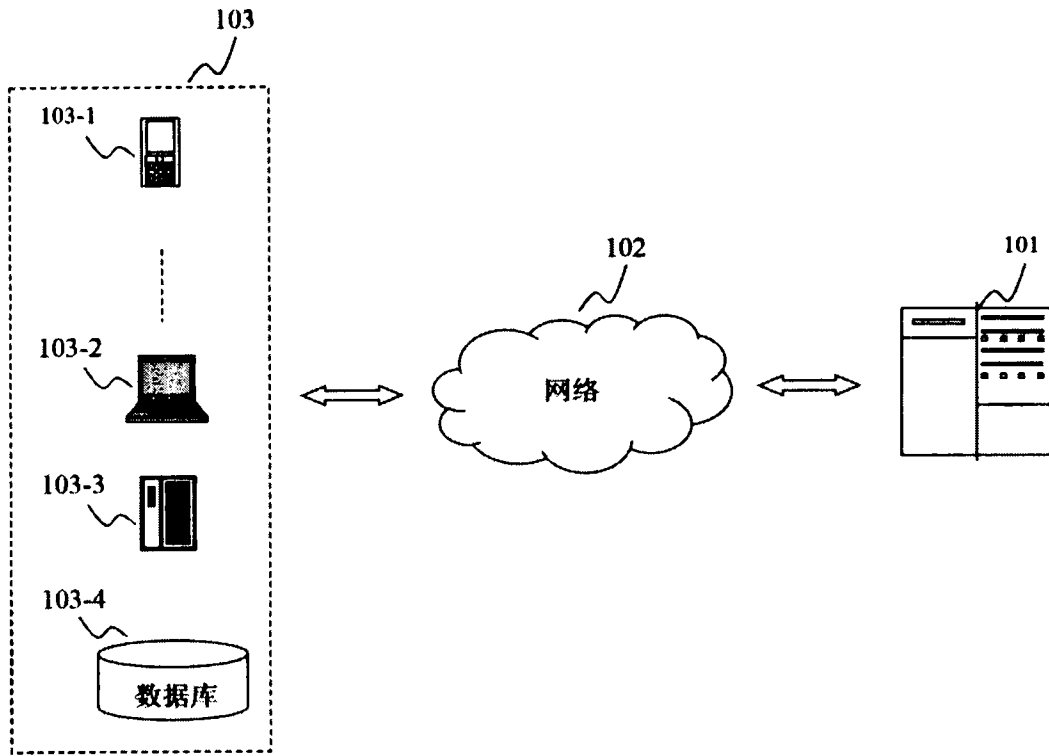


图 1

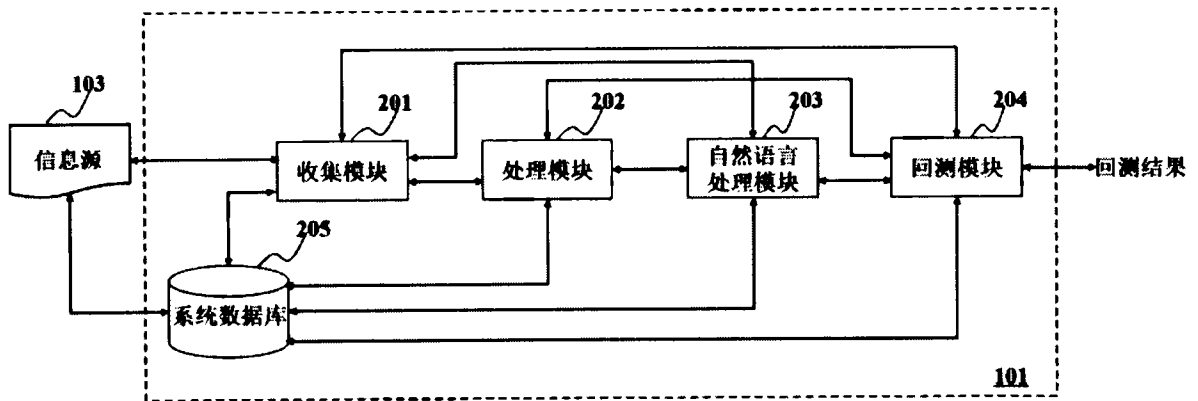


图 2

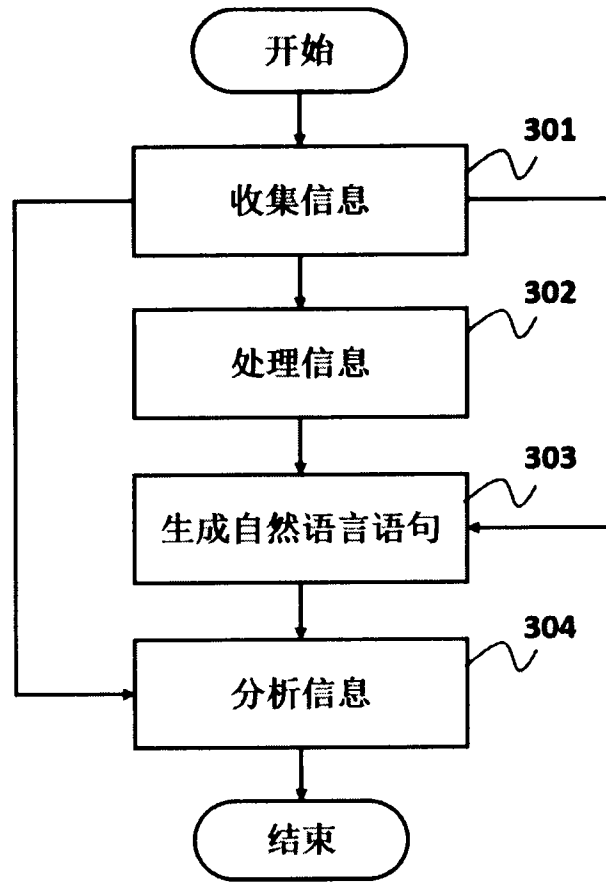


图 3

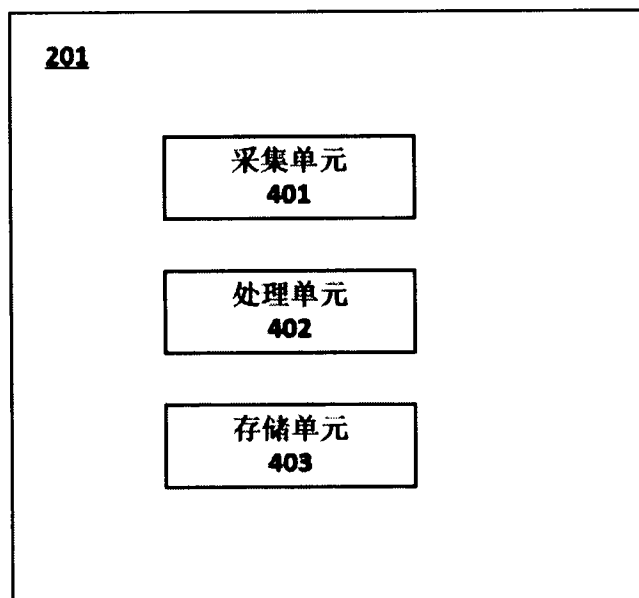


图 4

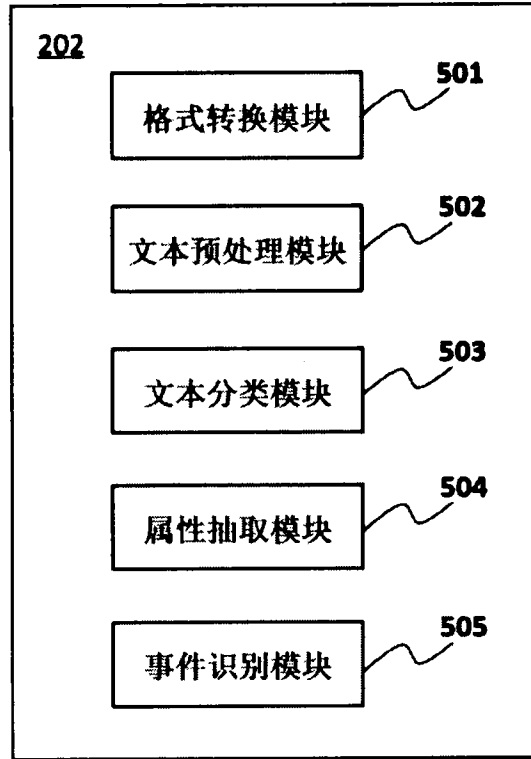


图 5

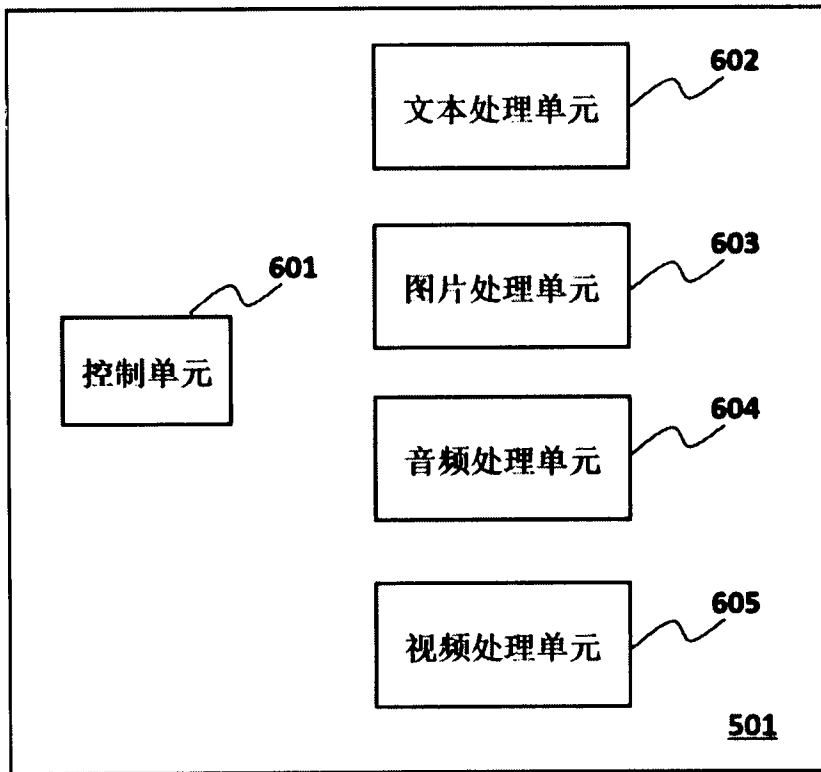


图 6

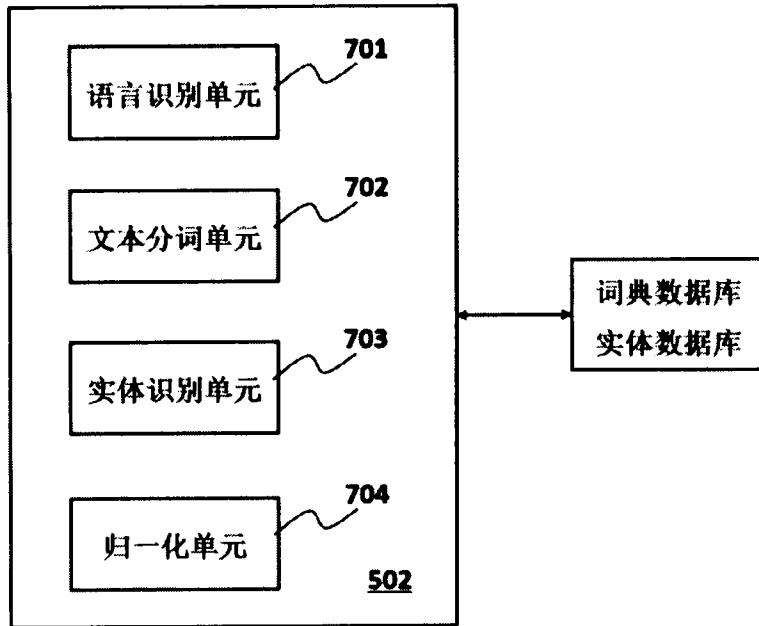


图 7

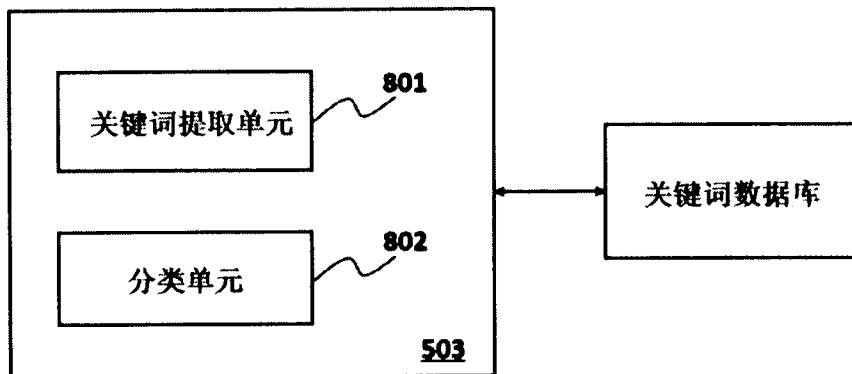


图 8

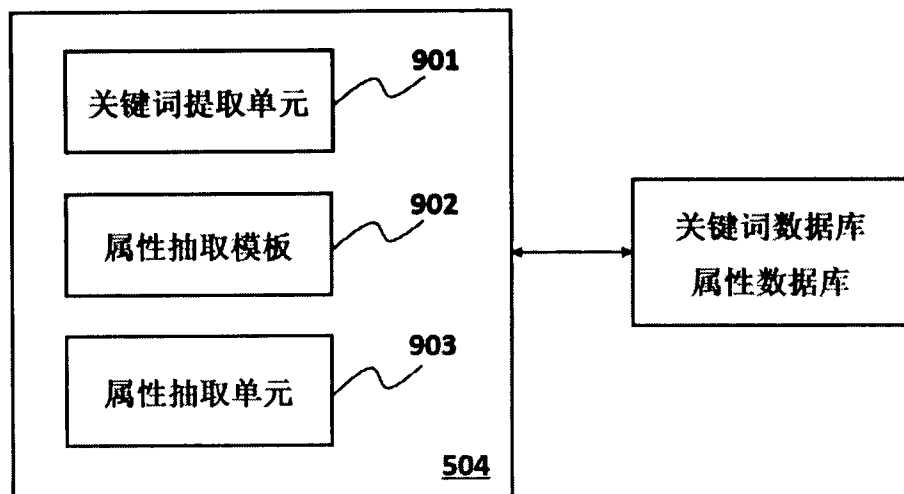


图 9



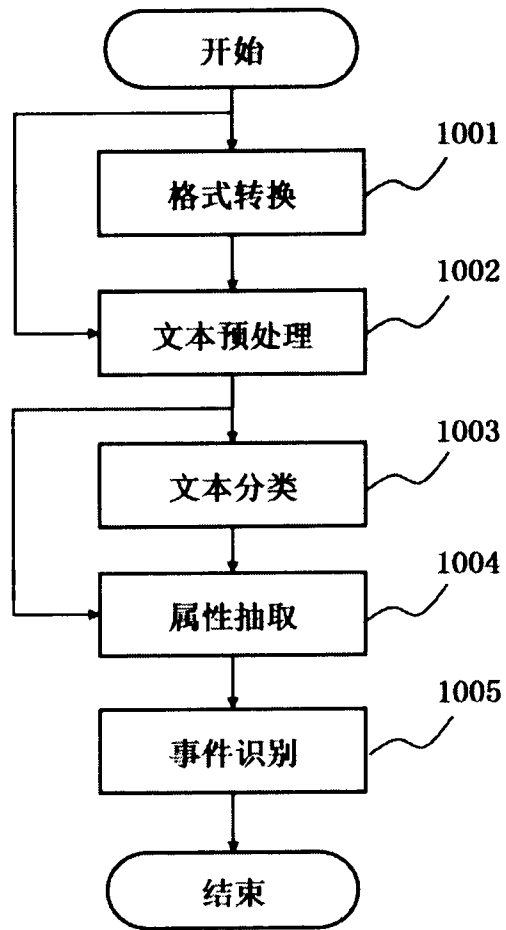


图 10

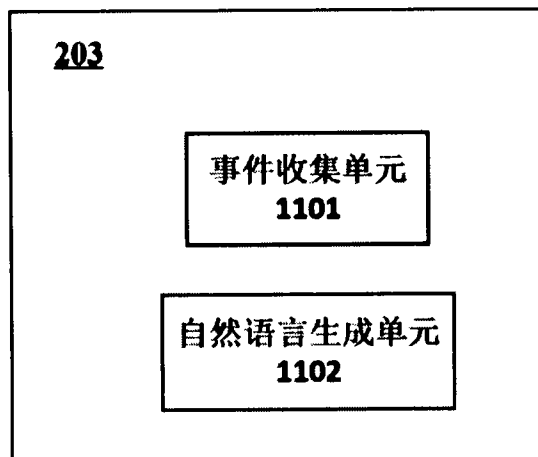


图 11

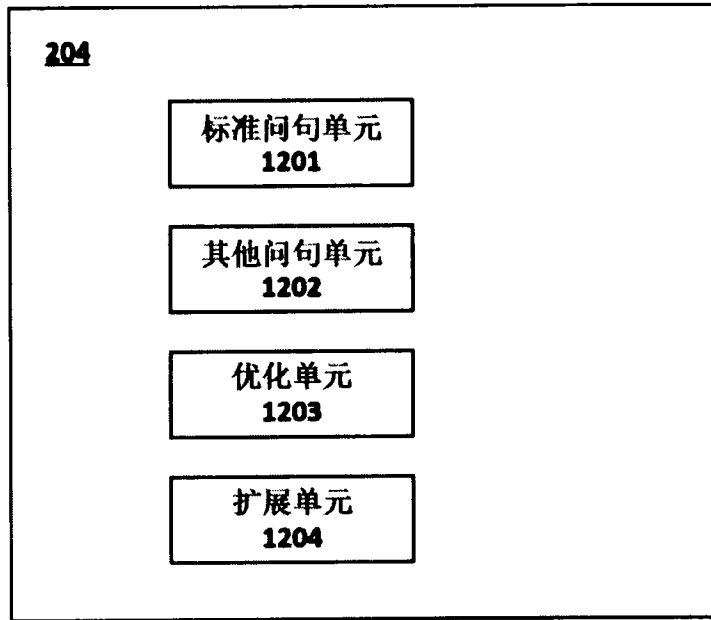


图 12

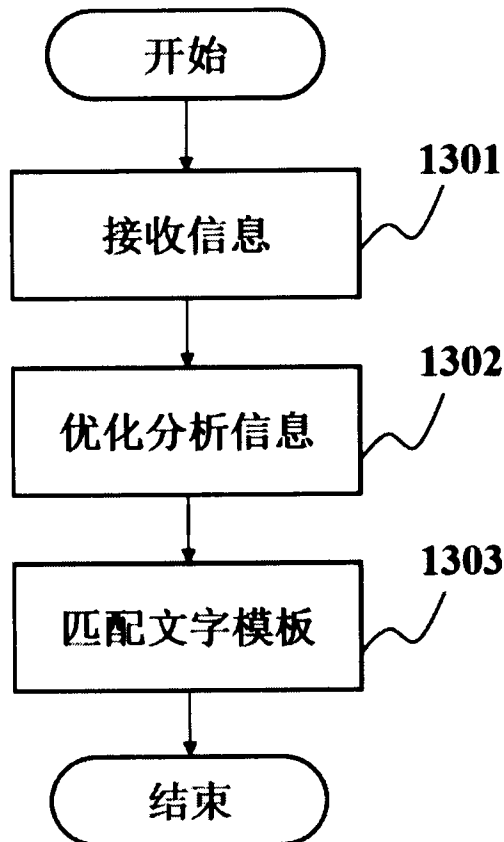


图 13

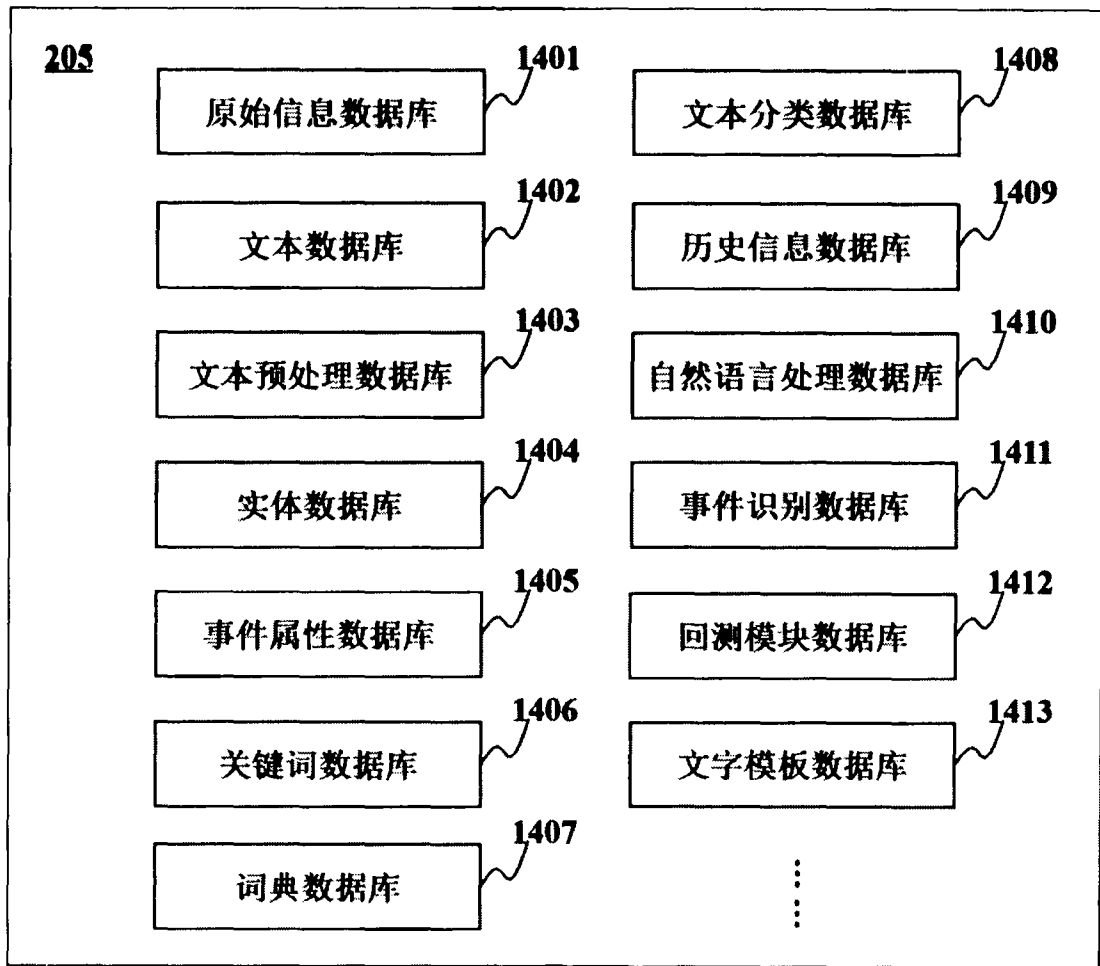


图 14

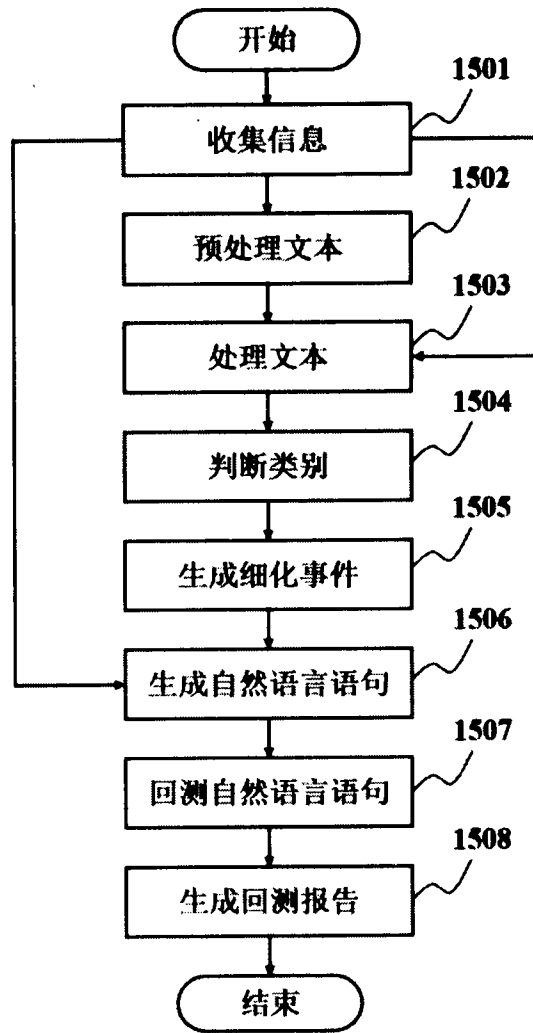


图 15

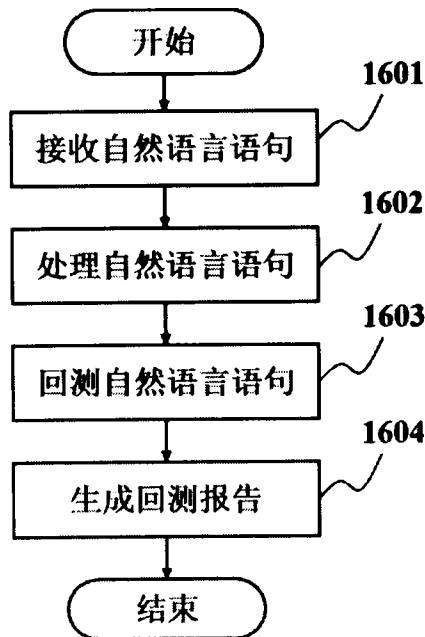


图 16

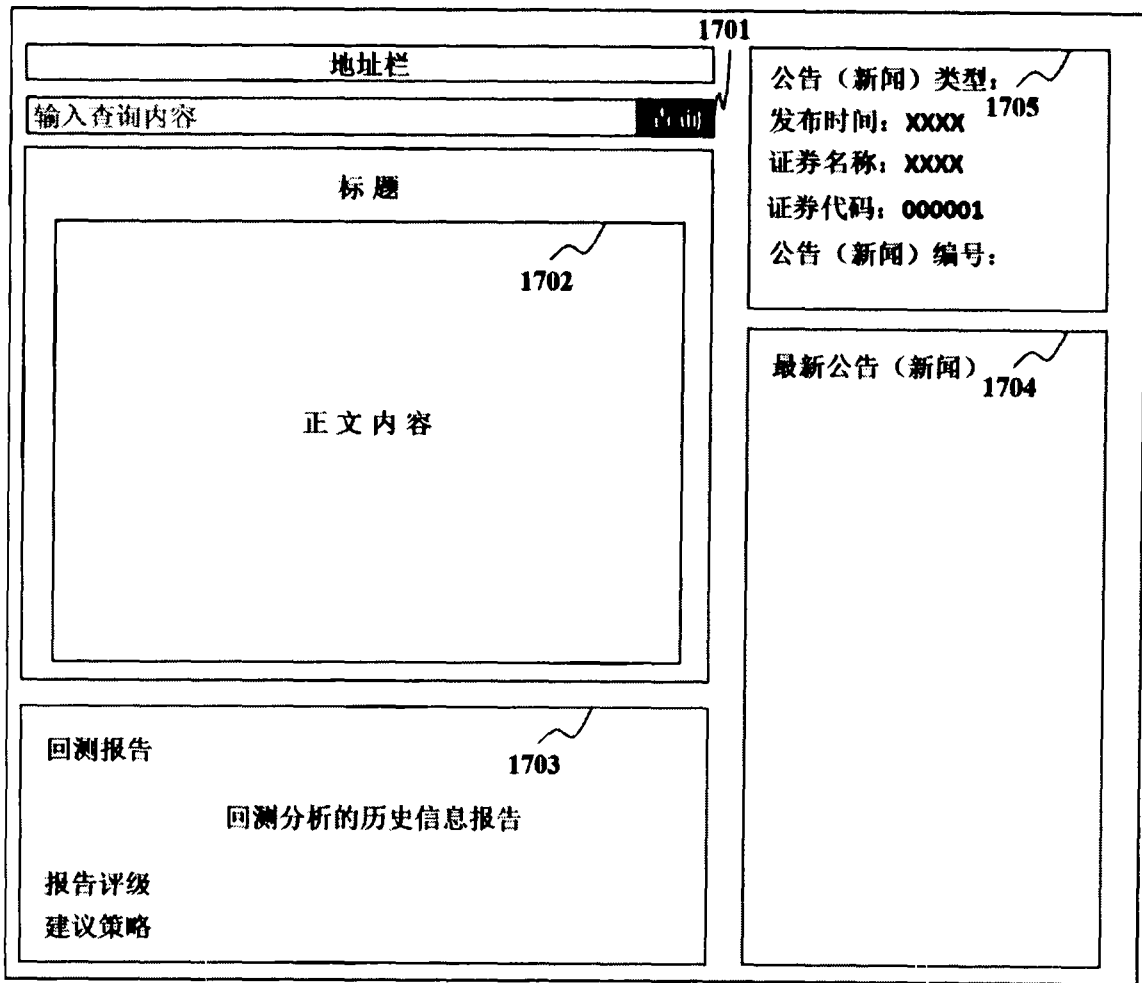


图 17

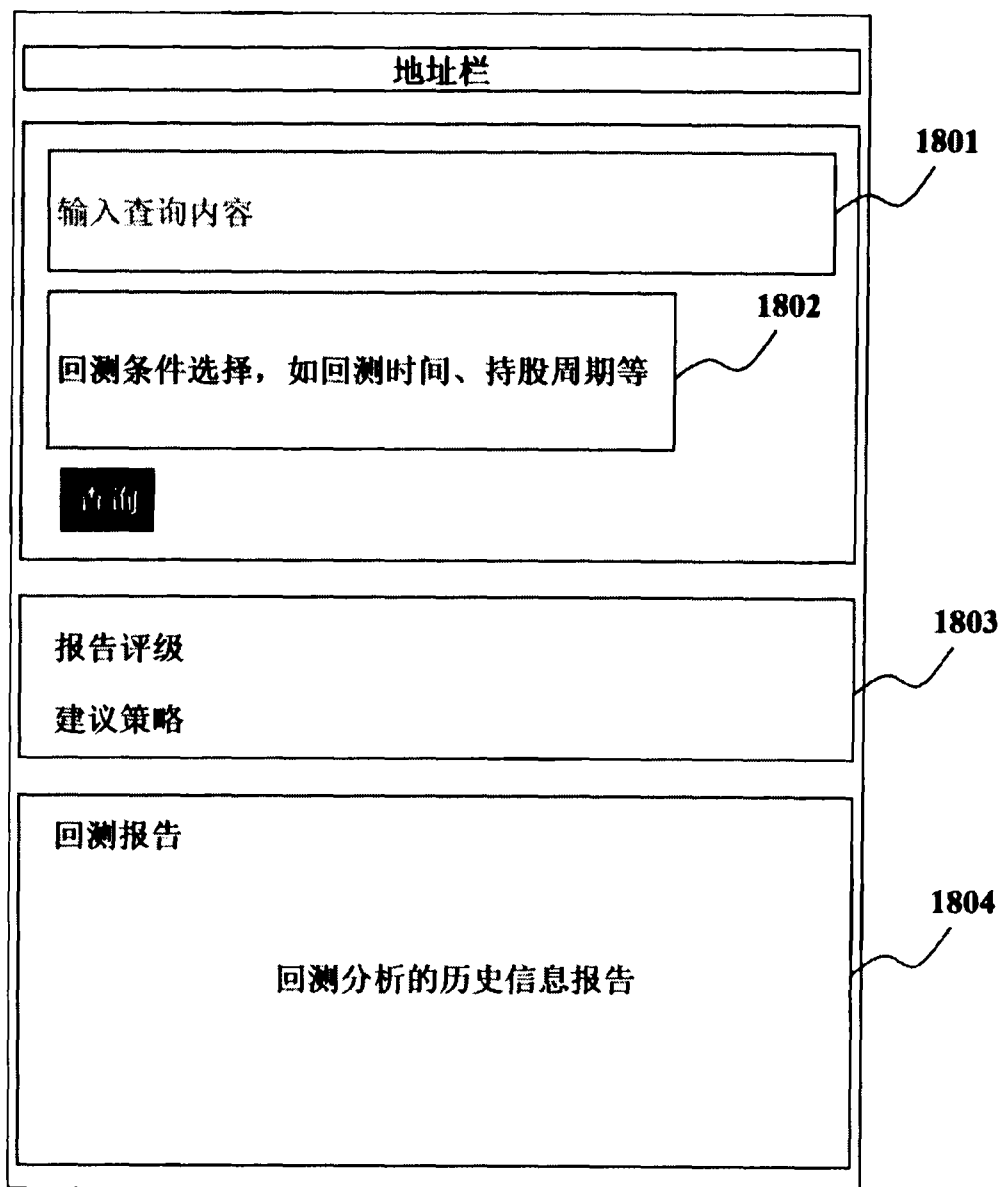


图 18



中信证券  
CITIC SECURITIES

中信证券股份有限公司 2014 年年度报告

公司简称：中信证券

中信证券股份有限公司

2014 年年度报告

第三节 会计数据和财务指标摘要

一、近三年主要会计数据和财务指标  
(一) 主要会计数据

单位：元 币种：人民币

13/251

中信证券股份有限公司 2014 年年度报告

项目	2014 年度	2013 年度	本期末比上年同期增减 (%)	2012 年度
营业收入	29,197,531,133.19	16,115,272,156.70	81.18	11,693,881,925.78
归属于母公司股东的净利润	11,371,193,825.46	5,243,916,979.11	116.20	4,237,418,476.37
归属于母公司股东的扣除非经常性损益的净利润	9,658,872,150.82	5,267,620,120.39	83.36	4,190,666,923.74
经营活动产生的现金流量净额	30,431,948,209.43	-18,609,569,343.95	不适用	-19,102,688,007.45
项目	2014 年 12 月 31 日	2013 年 12 月 31 日	本期末比上年同期增减 (%)	2012 年 12 月 31 日
归属于母公司股东权益	99,098,670,266.70	87,688,484,740.93	13.01	86,465,083,750.73
资产总额	479,626,450,225.21	271,354,248,951.82	76.75	168,507,554,655.12
负债总额	378,494,965,143.91	181,952,153,799.06	108.02	81,823,257,968.16
股本权益总额	101,131,485,081.30	89,402,095,152.76	13.12	86,684,296,686.96
其他综合权益	820,873,310.39	-1,226,048,782.49	不适用	-596,813,465.30

(二) 主要财务指标

项目	2014 年度或 2014 年 12 月 31 日	2013 年度或 2013 年 12 月 31 日	本期末/本期末比上年同期增减 (%)	2012 年度或 2012 年 12 月 31 日
基本每股收益 (人民币/股)	1.03	0.48	114.58	0.38
稀释每股收益 (人民币/股)	1.03	0.48	114.58	0.38
扣除非经常性损益后的基本每股收益 (人民币/股)	0.88	0.48	83.33	0.38
加权平均净资产收益率 (%)	12.18	6.02	增加 6.16 个百分点	4.90
扣除非经常性损益后的加权平均净资产收益率 (%)	10.38	6.05	增加 4.33 个百分点	4.84
归属于母公司股东的每股净资产 (人民币/股)	9.00	7.96	13.07	7.85
资产负债率 (%)	73.23	60.47	增加 12.76 个百分点	35.17

(三) 母公司财务报表及主要风险指标摘要

2014 年 12 月 31 日母公司净资产为人民币 443.19 亿元，同比增长 27.37%，主要由于报告期内公司发行次级债、净利润大幅增长所致。

项目	2014 年 12 月 31 日	2013 年 12 月 31 日
净资产 (人民币元)	44,319,246,529.25	34,796,488,465.78

图 19

- 重要提示**
- 一、本公司董事会、监事会及董事、监事、高级管理人员保证年度报告内容的真实、准确、完整，不存在虚假记载、误导性陈述或重大遗漏，并承担个别和连带的法律责任。
  - 二、本报告经本公司第五届董事会第三十五次会议审议通过。本公司全体董事出席董事会会议，未有董事对本报告提出异议。
  - 三、安永华明会计师事务所（特殊普通合伙）为公司出具了标准无保留意见的审计报告。
  - 四、公司负责人王东明先生、主管会计工作负责人葛小波先生及会计机构负责人葛小波先生声明：保证年度报告中财务报告的真实、准确、完整。
  - 五、本公司经董事会审议的 2014 年度利润分配预案为：以本公司截至 2014 年 12 月 31 日的总股数计算，每 10 股派发现金红利人民币 3.10 元（含税）；如本公司于 2014 年度 H 股股息派发基准日时已完成新增发行 H 股事宜，2014 年度每股派发现金红利的金额将在人民币 3,415,241,604.00 元（含税）的范围内，以发行后的总股数为基数做相应调整。此预案尚需公司股东大会批准。
  - 六、本报告所涉及的未来计划、发展战略等前瞻性描述不构成本公司对投资者的实质承诺，敬请投资者注意投资风险。
  - 七、是否存在被大股东及其关联方/连方非经营性占用资金情况：否
  - 八、是否存在违反规定决策程序对外提供担保的情况：否

1/251

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2015/098086**

## A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F; H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNTXT, CNKI, VEN, USTXT, WOTXT, EPTXT: document?, news, event?, format, classif+, natur+ w language, sentence?, abstract?, digest+, summar+, topic+, histor+, associat+, back+, semantic

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103473263 A (DALIAN UNIVERSITY OF TECHNOLOGY), 25 December 2013 (25.12.2013), description, paragraphs [0002]-[0023] and [0031]-[0069], and figures 1-3	1-18
X	CN 103488663 A (INTERNATIONAL BUSINESS MACHINES CORPORATION), 01 January 2014 (01.01.2014), description, paragraphs [0054]-[0101], and figures 1-3 and 6-8	1-18
X	US 2012240020 A1 (MCKEOWN, K.R. et al.), 20 September 2012 (20.09.2012), description, paragraphs [0025]-[0071], and figures 1-5	1-18
A	US 2015339269 A1 (KONCHITSKY, A.), 26 November 2015 (26.11.2015), the whole document	1-18

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>“&amp;” document member of the same patent family</p>
---	---

Date of the actual completion of the international search  
19 August 2016 (19.08.2016)

Date of mailing of the international search report  
**26 September 2016 (26.09.2016)**

Name and mailing address of the ISA/CN:  
State Intellectual Property Office of the P. R. China  
No. 6, Xitucheng Road, Jimenqiao  
Haidian District, Beijing 100088, China  
Facsimile No.: (86-10) 62019451

Authorized officer  
**PANG, Yan**  
Telephone No.: (86-10) **62089462**



**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
**PCT/CN2015/098086**

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103473263 A	25 December 2013	None	
CN 103488663 A	01 January 2014	US 2013332450 A1	12 December 2013
		US 2014195884 A1	10 July 2014
		DE 102013205737 A1	12 December 2013
US 2012240020 A1	20 September 2012	US 2012240032 A1	20 September 2012
		US 2005203970 A1	15 September 2005
		CA 2496567 A1	25 March 2004
		WO 2004025490 A1	25 March 2004
		US 8176418 B2	08 May 2012
US 2015339269 A1	26 November 2015	None	

国际检索报告

国际申请号

PCT/CN2015/098086

<p>A. 主题的分类</p> <p>G06F 17/30(2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																										
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06F; H04L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CPRSABS;CNTXT;CNKI;VEN;USTXT;WOTXT;EPTXT:文件, 新闻, 事件, 格式, 分类, 自然语言, 语句, 摘要, 总结, 主题, 历史, 关联, 回溯, 回溯, 语义, document?, news, event?, format, classif+, natur+ w language, sentence?, abstract?, digest+, summar+, topic+, histor+, associat+, back+, semantic</p>																										
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 103473263 A (大连理工大学) 2013年 12月 25日 (2013 - 12 - 25) 说明书第[0002]-[0023]、[0031]-[0069]段, 图1-3</td> <td>1-18</td> </tr> <tr> <td>X</td> <td>CN 103488663 A (国际商业机器公司) 2014年 1月 1日 (2014 - 01 - 01) 说明书第[0054]-[0101]段, 图1-3、6-8</td> <td>1-18</td> </tr> <tr> <td>X</td> <td>US 2012240020 A1 (MCKEOWN KATHLEEN R等) 2012年 9月 20日 (2012 - 09 - 20) 说明书第[0025]-[0071]段, 图1-5</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>US 2015339269 A1 (ALON KONCHITSKY) 2015年 11月 26日 (2015 - 11 - 26) 全文</td> <td>1-18</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型:          “A” 认为不特别相关的表示了现有技术一般状态的文件          “E” 在国际申请日的当天或之后公布的在先申请或专利          “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)          “O” 涉及口头公开、使用、展览或其他方式公开的文件          “P” 公布日先于国际申请日但迟于所要求的优先权日的文件          “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件          “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性          “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性          “&amp;” 同族专利的文件</p> <table border="1"> <tr> <td>国际检索实际完成的日期</td> <td>国际检索报告邮寄日期</td> </tr> <tr> <td>2016年 8月 19日</td> <td>2016年 9月 26日</td> </tr> <tr> <td>ISA/CN的名称和邮寄地址</td> <td>受权官员</td> </tr> <tr> <td>中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</td> <td>庞艳</td> </tr> <tr> <td>传真号 (86-10)62019451</td> <td>电话号码 (86-10)62089462</td> </tr> </table>		类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 103473263 A (大连理工大学) 2013年 12月 25日 (2013 - 12 - 25) 说明书第[0002]-[0023]、[0031]-[0069]段, 图1-3	1-18	X	CN 103488663 A (国际商业机器公司) 2014年 1月 1日 (2014 - 01 - 01) 说明书第[0054]-[0101]段, 图1-3、6-8	1-18	X	US 2012240020 A1 (MCKEOWN KATHLEEN R等) 2012年 9月 20日 (2012 - 09 - 20) 说明书第[0025]-[0071]段, 图1-5	1-18	A	US 2015339269 A1 (ALON KONCHITSKY) 2015年 11月 26日 (2015 - 11 - 26) 全文	1-18	国际检索实际完成的日期	国际检索报告邮寄日期	2016年 8月 19日	2016年 9月 26日	ISA/CN的名称和邮寄地址	受权官员	中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	庞艳	传真号 (86-10)62019451	电话号码 (86-10)62089462
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 103473263 A (大连理工大学) 2013年 12月 25日 (2013 - 12 - 25) 说明书第[0002]-[0023]、[0031]-[0069]段, 图1-3	1-18																								
X	CN 103488663 A (国际商业机器公司) 2014年 1月 1日 (2014 - 01 - 01) 说明书第[0054]-[0101]段, 图1-3、6-8	1-18																								
X	US 2012240020 A1 (MCKEOWN KATHLEEN R等) 2012年 9月 20日 (2012 - 09 - 20) 说明书第[0025]-[0071]段, 图1-5	1-18																								
A	US 2015339269 A1 (ALON KONCHITSKY) 2015年 11月 26日 (2015 - 11 - 26) 全文	1-18																								
国际检索实际完成的日期	国际检索报告邮寄日期																									
2016年 8月 19日	2016年 9月 26日																									
ISA/CN的名称和邮寄地址	受权官员																									
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	庞艳																									
传真号 (86-10)62019451	电话号码 (86-10)62089462																									

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2015/098086

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103473263	A	2013年 12月 25日	无			
CN	103488663	A	2014年 1月 1日	US	2013332450	A1	2013年 12月 12日
				US	2014195884	A1	2014年 7月 10日
				DE	102013205737	A1	2013年 12月 12日
US	2012240020	A1	2012年 9月 20日	US	2012240032	A1	2012年 9月 20日
				US	2005203970	A1	2005年 9月 15日
				CA	2496567	A1	2004年 3月 25日
				WO	2004025490	A1	2004年 3月 25日
				US	8176418	B2	2012年 5月 8日
US	2015339269	A1	2015年 11月 26日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)