



(12) 发明专利申请

(10) 申请公布号 CN 104471571 A

(43) 申请公布日 2015.03.25

(21) 申请号 201380037182.3

(74) 专利代理机构 北京律智知识产权代理有限公司 11438

(22) 申请日 2013.07.11

代理人 姜燕 王卫忠

(30) 优先权数据

61/670,481 2012.07.11 US

(51) Int. Cl.

G06F 17/30(2006.01)

(85) PCT国际申请进入国家阶段日

2015.01.12

(86) PCT国际申请的申请数据

PCT/CN2013/079215 2013.07.11

(87) PCT国际申请的公布数据

W02014/008866 EN 2014.01.16

(71) 申请人 谢晚霞

地址 210014 江苏省南京市秦淮区永智路6号中国云计算创新基地B座205

(72) 发明人 谢晚霞

权利要求书2页 说明书13页 附图5页

(54) 发明名称

事件驱动架构下对Web活动索引、排序和分析的系统和方法

(57) 摘要

公开了一种用于组织Web活动的系统，包括：解析模块，用于接收和解析所述Web活动；概念索引模块，用于根据概念索引中的多个概念来索引所述Web活动；Web事件创建模块，用于从所述Web活动中生成多个Web事件；Web活动索引模块，用于根据Web事件索引中的所述多个Web事件来索引所述Web活动；概念代码管理模块，用于生成多个概念代码，每个概念代码分别与所述多个概念中的至少一个相关联；以及数据库，用于存储所述概念索引、所述Web事件索引和所述多个概念代码。



1. 一种用于组织 Web 活动的系统,包括 :

解析模块,用于接收所述 Web 活动 ;

概念索引模块,用于根据概念索引中的多个概念来索引所述 Web 活动 ;

Web 事件创建模块,用于从所述 Web 活动中生成多个 Web 事件 ;

Web 活动索引模块,用于根据 Web 事件索引中的所述多个 Web 事件来索引所述 Web 活动 ;

概念代码管理模块,用于生成多个概念代码,每个概念代码分别与所述多个概念中的至少一个相关联 ;以及

数据库,用于存储所述概念索引、所述 Web 事件索引和所述多个概念代码。

2. 根据权利要求 1 所述的系统,还包括概念创建模块,用于从所述 Web 活动中生成所述多个概念。

3. 根据权利要求 2 所述的系统,其中所述概念创建模块包括 :

语义模块 ;

情绪模块 ;以及

分类模块。

4. 根据权利要求 1 所述的系统,还包括社交图谱分析模块,用于分析社交网络。

5. 根据权利要求 1 所述的系统,还包括影响者排序模块,用于确定所述 Web 活动的创建者的影响力。

6. 根据权利要求 1 所述的系统,还包括概念代码的信息补充和丰富的模块。

7. 根据权利要求 1 所述的系统,还包括 :

Web 事件捆绑关联模块 ;以及

Web 活动和 Web 事件的描述生成模块。

8. 根据权利要求 1 所述的系统,还包括 API,用于与外部应用交互。

9. 一种用于组织 Web 活动的方法,包括 :

接收所述 Web 活动 ;

解析所述 Web 活动 ;

根据概念索引中的多个概念索引所述 Web 活动 ;

从所述 Web 活动中生成多个 Web 事件 ;

根据 Web 事件索引中的所述多个 Web 事件索引所述 Web 活动 ;

生成多个概念代码,其中每个概念代码分别与所述多个概念中的至少一个相关联 ;以及

将所述概念索引、所述 Web 事件索引和所述多个概念代码存储在数据库中。

10. 根据权利要求 9 所述的方法,还包括从所述 Web 活动中生成所述多个概念。

11. 根据权利要求 10 所述的方法,其中所述从所述 Web 活动中生成所述多个概念包括 :

对所述 Web 活动进行语义分析 ;

确定所述 Web 活动的情绪信息 ;

确定所述 Web 活动的权威性 ;以及

基于特定的分类架构体系来确定所述 Web 活动的类别。

12. 根据权利要求 9 所述的方法,还包括:

识别所述 Web 活动内的第一个 Web 参与者;

确定所述第一个 Web 参与者与第二个 Web 参与者在社交网络里的关系;

根据所述关系生成所述多个 Web 事件中的至少一个。

13. 根据权利要求 9 所述的方法,还包括确定所述 Web 活动的创建者的影响力。

14. 根据权利要求 9 所述的方法,还包括对所述多个概念代码中的一个进行信息补充和丰富。

15. 根据权利要求 9 所述的方法,还包括:

捆绑关联所述多个 Web 事件中的第一个 Web 事件和第二个 Web 事件;以及

生成所述 Web 活动、所述第一个 Web 事件和所述第二个 Web 事件的描述。

16. 根据权利要求 9 所述的方法,还包括与 API 交互。

17. 一种用于组织 Web 活动的系统,包括:

监视模块,用于侦测 Web 活动;

解析模块,用于接收所述 Web 活动;

概念创建模块,用于从所述 Web 活动中生成多个概念;

概念索引模块,用于根据概念索引中的所述多个概念来索引所述 Web 活动;

Web 事件创建模块,用于从所述 Web 活动中生成多个 Web 事件;

Web 活动索引模块,用于根据 Web 事件索引中的所述多个 Web 事件来索引所述 Web 活动;

概念代码管理模块,用于生成多个概念代码,其中每个概念代码分别与所述多个概念中的至少一个相关联;以及

数据库,用于存储所述概念索引、所述 Web 事件索引和所述多个概念代码。

18. 一种用于组织 Web 活动的方法,包括:

检测 Web 活动;

解析所述 Web 活动;

从所述 Web 活动中生成多个概念;

根据概念索引中的所述多个概念索引所述 Web 活动;

从所述 Web 活动中生成多个 Web 事件;

根据 Web 事件索引中的所述多个 Web 事件索引所述 Web 活动;

生成多个概念代码,每个概念代码分别与所述多个概念中的至少一个相关联;以及

将所述概念索引、所述 Web 事件索引和所述多个概念代码存储在数据库中。

事件驱动架构下对 Web 活动索引、排序和分析的系统和方法

[0001] 优先权 / 临时申请的交叉引用

[0002] 本申请要求享有在 2012 年 7 月 11 日提交的美国临时申请号 61/670,481 的优先权，该申请的全部内容被参考合并于此。

技术领域

[0003] 本发明的实施方案涉及一种用于分析互联网上的信息内容的系统和方法。更具体地来说，是关于一种用于对互联网内容进行索引和排序的系统和方法。虽然本发明的实施方案的应用很广泛，但尤其适用于将传统的互联网内容与诸如移动应用、社交媒体、众包媒体 (crowd sourced media)、以及博客之类的新媒体内容融合的应用。

背景技术

[0004] 总的来说，自从 Web 浏览器诞生以来，让用户在互联网上有效地浏览、发现、过滤以及参与一直是一个挑战。以高效的方式找到及时并且相关的信息是所有互联网用户的目标。考虑到内容构成的动态性，以及内容来源定义的多样性，实现这个目标尤其具有挑战性。过去，在线的内容主要由网站发布者在网站上发布，而现在，这个格局已经发生变化，许多在线内容是通过博客、微博、视频、图像、评论、用户评价以及社交网络发布。在移动设备上产生的内容和活动变得越来越多。举例来说，社交网络的内容包括状态更新、推文 (tweet)，转发 (re-tweet)、微博以及用户行为，例如赞 (like)、签到、书签、钉选 (pin) 和收藏。

[0005] 在过去的十来年中，Web 用户在 Web 上导航的主要模型为搜索引擎模型。当前的各种技术实施方式依赖于很多方法以便将相关内容提供给用户，但决定相关性的最重要因素仍然是外部链接（参见例如美国专利号 6,285,999）和关键字索引。这些技术手段之所以有效，是因其捕捉到了那个时候的主要的用户行为活动，即添加链接指向其他网站和点击链接的行为。这种倚重于外部链接和关键字索引的技术解决方式的结果是一种利用众包方式来决定信息相关性的模型，其实质上是人气竞赛。然而，这种模型的优势同时也是其最大的弱点，该弱点是过于关注网页和基于文本的内容。随着各种新内容形式的出现，和线上影响力评估的日益流行，这种方法已经不再适用，因其不能捕捉这种新的信息。随着在线用户行为和活动的巨幅增长，如上所述，外部链接和点击次数这两个维度过于简单化，无法体现新的 Web 活动的复杂性。其结果是大量有价值的信息丢失，导致在线用户的信息访问行为受挫和效率低下。

[0006] 例如，当今的搜索引擎没有支持捕捉用户行为、参与的用户、用户间的信息流通以及其他类型的 Web 活动（不同于点击次数和链接）的框架。此外，由于在对影响力的判定上，此类搜索引擎倚赖基于外部链接的人气竞赛，所以其带有历史偏见性。在这种模型中，一个内容相关性较强的网站如想获得很多外部链接，特别是在涉及到热门的搜索关键字的情况下，需要等待大量时间。正因如此，当前的搜索引擎工作模式是一种向后看的滞后模

式,最适宜于确定内容的过去相关性,但却不适宜于判定那些较新的,尚未流行的内容的相关性。

[0007] 当相同的内容出现在多个数据来源中时也会发生问题,这是很常见的情况。一些数据来源可能频繁地更新,而有些数据来源可能根本不会更新。因此,当信息在一个数据来源首先被更新时,最新的且准确的信息占少数。而众包方法却会给那些陈旧的信息更高的排名,因为它们被大多数其他数据来源所认可。在这些数据来源上的信息更新状况反映了那些藏在背后的隐含的行为。监视在不同数据来源上的信息更新情况可以用于对新的且准确的信息进行分析和排序。然而,搜索引擎和分析工具的当前实现方式忽略了这些隐含的行为,从而错过了可以用来对结果进行排序和分析的重要信号。

[0008] 此外,静态和动态网页的内容会随着时间的推移而被更新。但是目前的搜索系统是不考虑这点的,因其仅使用这些网页在某个时间点的内容快照。再者,线上内容不再整齐有序地存在于网页中,或者以纯文本的方式存在。因此,以网页链接和基于文本的关键字索引为重心的搜索引擎技术不再能够以最优的方式帮助用户寻找到相关内容。

[0009] 最近的一些技术发展(例如社交网络、博客、微博以及基于用户的行为系统)已经将互联网和移动互联网从一个基于文本文档的 Web 转变为一个行为和活动的 Web。创建这种新类型内容的基于行为的系统的例子包括内容积酿(curation)应用(如 Digg)、社交书签网站(如 Delicious 和 Pinterest)、转发应用(如 Tweetmeme)、共享平台(如 Twitter(推特)、微博和 Tumblr)、评论系统(如 Disqus 和 Echo)、基于位置应用的签到系统(如 Foursquare)等等。在 Web 上(以及在移动设备中)的用户行为和活动的数量由于这些新近出现的技术而大幅增加。与上述技术中的显式用户行为相比,网页(或应用等)随时间推移而产生的内容变化反映了后台的隐式用户行为。通过监视内容变化,可以在系统中捕捉这些隐式行为来进行智能分析。

[0010] 近几年来用户身份也收到更大的重视。Twitter(一个微博平台)已经围绕公开的用户资料和微消息建立了一个社区。评论系统如 Disqus 和 Echo 能够使用户以单一的身份(这个身份包括用户名和 / 或照片)对数以千计的博客进行评论。很多 Web 应用已经开始基于用户在 Twitter、LinkedIn 以及其它社交网络中所发布内容的访问流量和粉丝的数量来对其线上影响力进行衡量和评分。因此,虽然仅在几年以前,对线上影响力这一“货币”的衡量还只取决于网站的独立访客数和外部网站链接数量,现在,线上影响力的衡量还需要考虑用户自身的线上影响力。

[0011] 在实时搜索领域,一些新兴技术开始出现,试图来突破当前搜索引擎方法的局限性。通常,这些技术试图聚焦于那些正在流行的链接,链接的流行程度取决于它们在社交网络中被分享及转发的频率。这些方法有助解决直接相关性的问题,但是在分析和衡量主题相关性、主题内参与者、主题所涉人物之间的关系、人物和主题的关系、这些关系的变化、主题内所发生的活动类型等方面,仍然不足以提供一套全面的方法体系。对人气的聚焦注定了这些方法带有滞后性。此外,由于这些系统主要聚焦于那些能够便捷提供这些线上活动数据的平台(如 Twitter),它们仅仅捕捉了互联网上线上活动数据的一小部分。实质上,这些系统只不过是在老的方法体系中引入了一些小的改进而已,并未能真正捕捉到互联网上那些围绕线上内容(包括文档和基于行为的内容)、线上参与者、以及 Web 活动方面所出现的错综复杂的发展变化。

[0012] 其结果是，传统的基于 Web 的搜索和新兴的实时搜索都不能给用户提供足够的 Web 能见度，因其实施方式太过简单化，无法反映 Web 上新增的用户行为和活动的类型，以及相关的复杂性。两种实施方式都无法帮助用户获得关于那些在特定主题领域具有影响力线上参与者的数据和信息。相反地，这两种实施方式都仅关注指向 Web 内容的链接，而不是突出那些新的内容，即在这种新形态 Web 中创建了这些线上内容的用户。两种方式都不能有效地帮助用户及时发现 Web 上围绕用户所感兴趣的话题展开的那些，正在发生的而且大家积极参与的讨论，尽管这些讨论代表着一个非常丰富的线上内容来源。相反地，这两种实施方式都是基于外人不清楚的算法，输出一个黑盒式的网页链接列表（搜索结果列表）。总而言之，当前的这些实现方法都不能够由点及面地链接和分析信息，因此不能为用户提供一个在互联网上高效地探索、发现和积极参与的导航仪。其结果是 Web 用户沦为了网页快照的历史学家，不能获得足够的能见度，用户体验显著受挫。

[0013] 目前社交网络的实施方案确实提供了一个围绕人物和 Web 内容贡献者的具有吸引力的工具。在该框架中，用户可以参考其社交图谱中其他用户的推荐，对 Web 内容进行积酿（Curation）。但是当前的社交网络仅提供这一种方式的线上内容，并且局限在其隔离的有限空间内。例如，如果用户在 Twitter 上搜索，并不等同于搜索 Web。他所搜索到的仅仅是一小部分信息。例如，博客网站上用户之间的讨论和互动将不会被社交网络捕捉。如果用户仅依赖于其社交网络来采集 web 信息，由于其社交网络的局限性，将会显得“目光短浅”。由于重点关注 Web 参与者，和传统的搜索技术相比，当前的实施方案走到了另一个对立极端。其模型太过于以用户为中心，缺少一个能够智能地将用户生成的内容（即“User Generated Content”，简称为“UGC”）和其他类型的线上内容有效融合的框架体系。

[0014] 其结果是 Web 被割裂为两个阵营：专门管理和索引内容的阵营，和专门管理社交图谱的阵营。两者都不能全面捕捉在 Web 用户、网站、在线行为和在线内容（包括原始的和索引过的内容）之间存在的复杂的多级相互关联性。用户只能勉强借助于这两个分离的技术手段在 Web 上采集信息，造成效率极为低下、信息超载和令人沮丧的用户体验。

发明内容

[0015] 本发明的典型实施方案通过架构体系性改进，针对性地解决了以上所述的一些问题。本发明不仅解决了上述的一个或多个问题，而且提供了一个可以预测内容相关性的框架，从而帮助 Web 用户更快地发现那些对于他们有重要意义的信息，并更早参与到相关的对话讨论中。

[0016] 本发明的实施方案可以包含多个进程、模块或子系统，包括：实时抓取（crawling）和聚合子系统、推送信息（feed）处理子系统、解析子系统、社交图谱分析子系统、概念索引子系统、活动索引子系统、语义分析子系统、情绪分析（sentiment）子系统、分类子系统、影响力排序子系统、Web 事件创建子系统、Web 活动捆绑关联和管理子系统、概念代码（Ticker）管理和创建子系统、概念代码补充和丰富子系统、Web 活动和 Web 事件排序子系统、Web 活动和 Web 事件描述生成子系统、Web 信息流管理子系统、数据存储系统、开发者配置和管理子系统、事件路由（event-routing）分配子系统，用于过滤的基于规则的事件子系统、用于关联、分析及预测事件的复杂事件处理模块或子系统、认证子系统、Web 或移动应用、用以实施专有 Web 索引的一体机设备，以及 API。

[0017] 本发明实施方案可以聚合和索引 Web 活动。Web 活动可以包括公开的 Web 内容和私有 Web 内容,例如社交网络(如 Facebook(脸书)或 Twitter 或微博)中用户的私有信息流(private feed)。Web 活动还包括通过持续监测 Web 内容的更新、任何由用户或者应用在互联网上或移动装置(如手机)上所生成活动和行为而推断出来的隐性用户行为。Web 活动可以进一步包括公开或内部的数据记录(例如文件、电子邮件和实时消息),使用专有或第三方分析和算法工具所推演出的活动和属性,从第三方 API 获得的活动信息,存在于用户的社交图谱、内容、标签和元数据中的显性和隐性的活动及变化。

[0018] Web 活动的例子可以包括状态更新、推文(tweets)、转发(re-tweets)、微博、评论、签到、收藏、赞、踩(dislikes)、分享、钉选(pin)、新的概念和主题、新的 Web 参与者、社交网络和移动设备的应用商店中的应用下载、概念的相关活动水平及变化、Web 参与者的活动水平的变化、概念内新参与者的行、概念内参与者的重复行为、概念内用户的线上影响力、概念内用户的线上影响力的变化、用户对于概念的态度、用户对概念的态度的变化、信息在网站或应用之间的流动、信息在 Web 参与者之间的流动、内容的地理位置、内容在 Web 上出现的位置、内容在网页上的位置、内容类型(包括但不限于博客、图像、视频、评论和状态更新)、内容质量和分类(例如,垃圾内容或权威信息、内容语言分类)、信息在一段时间内的传播路径、参与者的 Web 行为发生的相对时间、点击率、用户的显式社交图的结构变化、用户的隐式社交图的结构变化(用户的隐式社交图是通过用户在 Web 对话中与其他用户的互动交流信息来生成)、用户的社交资料的变化、用户或用户的社交图引用的概念和主题的变化、Web 元数据、用户元数据、概念元数据、内容包含的情绪信息、概念的趋势、Web 活动的增量变化、以及内容与 Web 参与者之间出现的新关系。

[0019] 本发明实例可以监视在一段时间内某个具体数据源的 Web 内容的更新并从中获得隐含的行为活动。例如,一个企业或个人的联系信息可能出现在多个数据源上,这些数据源对这个联系信息可能进行不同的更新。本发明实例可以结合机器学习和聚类技术对不同数据源的更新活动进行分析,以从多个数据源中找到权威信息并发现隐含的模式和规律。

[0020] 本发明实例可以监视线上的内容和活动,通过一个被称为概念索引的过程来识别并记录 Web 上的概念。概念可以是在 Web 上出现的任意一组关键字,如本发明所定义的,其代表了一个唯一的主题。与自上而下驱动的分类架构体系不同,这些主题可以自发组织以便反映线上内容的变化,虽然这些机制在本发明都可以利用。一些概念的示例可以是“猪流感”、“实时搜索”、“巴拉克·奥巴马”以及“微软收购雅虎”。主题的字数没有限制。本发明可以应用语义分析、聚类分析和模糊匹配技术来提取主题,这个过程应该考虑到关键字的同义词和语义。这能够使诸如“购买”、“收购”、“兼并”之类的关键字被归为同一主题,使得概念不受具体的关键字的限制,从而更好地反映真正的含义。

[0021] 与关键字索引相比,概念索引可以开启很多不同的功能。因为其能让 Web 用户以类似用户在社交网络(比如 Twitter 或微博)里关注其他用户的方式时刻关注概念。例如,当关注概念时,用户可以及时查看与概念有关的内容流,与该概念有关的所有元数据,以及与该概念有关的所有相关 Web 活动。在本发明的一个典型实施方案中,上述的 Web 活动可以被索引到概念。例如,在每个概念里,本发明的典型实施方案可以监视活动水平、情绪、趋势、Web 参与者以及诸如 URL 之类的相关数据源。这允许 Web 上的概念被时刻监视和跟踪。在一个典型实施方案中,用户可以使用限定在某个概念内的热门话题和关键字,这与其他

提供广泛的、通用的趋势的替代方案截然不同。

[0022] 本发明可以为每个概念创建标签或“概念代码”。概念代码可以等同于可编程的主题标签，当然其可以比关键字反映显著更多信息。例如，概念代码可以包括 Web 活动的信息。这可以允许用户和开发者使用概念代码（概念代码包含的查询包括关键字和 Web 活动）来搜索过去的 Web 内容或者订阅未来的 Web 内容。例如，用户可以搜索“猪流感”，同时还可以说明内容类型（视频、图像、评论等）、内容来源、权威级别、内容的情绪、和 / 或内容分类（购物、健康等）。这种方式可以允许用户精确地找到想要的信息。在另一个示例中，在线旅游出版社可以订阅只反映正面情绪的用户对酒店的评论。在这种示例的实施方案中，类似开发者构建其自己的应用程序的机制，概念代码可以充当对（过去和未来的）Web 内容和 Web 活动的查询语言。这对程序员们的好处在于，他们不需要自己构建 Web 活动索引和分析框架，而可以通过一个典型的实施方案中的 API 来利用本发明的实施方案中的功能。

[0023] 在本发明的一个典型实施方案中，概念代码使用第三方数据来源来补充数据。这些数据来源包括但不限于人工精选和编辑的数据来源（例如维基百科和 Freebase）、结构化数据来源（例如 Wolfram）、以及用户定义的元数据（其中用户可以创建私有的和公共的内容分类和类别）。在一个用户定义的示例中，用户可以提供关键字标签和“Web 活动标签”来指示本发明的实施方案怎样去索引 Web 活动。用户定义的元数据可以限制在私下使用，例如在企业内，也可以开放公开使用。

[0024] 本发明的实施方案可以包含一个配置和管理子系统，以便开发者或组织机构使用概念代码和 Web 事件来构建应用。在一个典型的实施方案中，本发明可以包括一个图形用户界面（GUI），以便开发人员轻松地构建概念代码以及访问本发明中的数据。

[0025] 在一个典型的实施方案中，所有 Web 活动使用一个专有的数据模型进行标准化和索引，以便发现和分析独特的相互关系。在一个典型的实施方案中，数据模型可以在关键字和概念之间创建相互关系，然后可以在概念、Web 参与者（例如人）、数据记录（例如 URL 或推文或微博）、前面每个元素的属性、以及派生属性之间创建相互关系。派生属性可以包括 Web 事件或者关于所存储的数据的任何分析结果。一个计算和存储派生属性的例子是投资银行针对期权使用其自己的专有期权定价模型来定期记录和存储期权的风险指标：delta 值、gamma 值、以及 theta 值。

[0026] 数据模型的结果可以生成概念、元数据和数据记录（例如 Web 链接）的独特的社交关系图。例如，每个概念可以与 Web 参与者和 URL 产生相互关系。或者，每个 Web 参与者可以与概念和 URL 产生相互关系。最后，每个 URL 可以与 Web 参与者和概念产生相互关系。由于概念包含 Web 活动，这可以超越基于关键字的方法来访问 Web 上的信息。取而代之的是，本发明可以允许用户通过下面的典型查询，如关键字、概念、Web 参与者、数据记录、元数据，或前面元素的任意组合，来查询 Web。

[0027] 本发明的实施方案使用例如事件处理和监视的框架和架构将索引的 Web 活动转变为 Web 事件。作为一个示例来说，博客上的一条用户评论可以被认为是一个 Web 活动。本发明可以像监视飞机的飞行路径和高度的雷达一样从这单一的 Web 活动中监视和识别几个 Web 事件。例如，根据一条用户评论这样的一个 Web 活动，以下的典型 Web 事件可以在本发明中被监视和记录：一个博客的一条新评论、从这个用户的评论中发现的新概念、以及一个概念中新的 Web 参与者。用这种方式，一个 Web 上的基本活动可以分解为能够被记录和

分析的许多事件。Web 事件可以包含时间戳信息，使得 Web 活动可以记录发生的 Web 事件时间序列。Web 事件可以存储在数据库中，并且在某些情况下，可以同时发送到内部和外部的订阅的应用和数据库。在一个典型的实施方案中，一个 Web 事件可以是基于事件的框架中的一个事件，与传统的事件有所区别的是，这里的每个事件都与一种具体类型的 Web 活动相关联。在本发明的一个典型实施方案中，Web 活动和 Web 事件可以被全部回放或者在概念中回放，使得用户可以看见这些事件是如何在 Web 上发展和发生的。

[0028] 在本发明的一个典型实施方案中，Web 或移动应用可以为用户提供一个 Web 的动态目录，其中这些相互关系被实时反映。这可以帮助用户了解 Web 是如何将内容、人物和网页链接相关联的。并且，此 Web 或移动应用可以生成活动热点图，这些热点图也可以限制为特定的概念或主题生成。

[0029] 在本发明的一个典型实施方案中，一旦 Web 活动被转化为 Web 事件，这些事件可以被智能地分析以及关联。复杂事件处理技术和量化算法可以被应用来处理这些事件，以预测相关性和未来的 Web 活动。在一个典型实施方案中，本发明可以将 Web 上的活动转变为可量化的事件后进行分析，分析方式就如同在金融市场中的算法交易 (Algorithm Trading) 或者政府的反恐情报分析工作中应用特定算法一样。在一个典型实施方案中，为了预测，例如，新 Web 参与者、有用的内容、或者新的内容源的渐渐增强的相关性，本发明可以在信息跨 Web 参与者传播的时候对跨 Web 参与者或跨数据来源的信息传播路径进行关联分析。用这种方式，本发明的实施方案可以具有前瞻性，这种方式与仅仅为用户提供历史相关性的方法截然不同。

[0030] 在本发明的一个典型实施方案中，本发明可以捆绑关联 Web 活动以形成自己的智能活动和事件。这样做的目的在于为用户提供一个独特的互联网活动快照，而不会为用户造成信息过载的负担。在一个典型实施方案中，本发明可以将概念内的活动和事件捆绑关联，使得用户能够快速了解主题的相关信息情报和活动。在另一个典型实施方案中，本发明通常可以捆绑关联活动和事件。信息情报的例子可以包括推荐（关于内容、数据来源、Web 参与者以及新的概念代码）；预测；凸显新概念以帮助用户发现；当概念的活动水平或 Web 参与者的活动水平的变化超过标准偏差时警告用户；根据在用户感兴趣的概念的相似概念内的 Web 参与者的影响力，向用户推荐其应当在其社交网络中关注的 Web 参与者；为用户建议有很多 Web 活动的 URL；基于用户订阅信息的建议；基于用户社交网络中隐含的活动、跟随者的情况以及其他 Web 活动如博客中的线上讨论对用户提出的建议。本发明可以允许用户在概念内指定其目标，使得系统可以为用户提供更加具体和个性化的信息情报。用户提供的目标的例子可以包括：营销、公共关系 (PR)、新的相关内容源、新的相关的人、竞争调查、或者产品研发。例如，如果用户选择营销作为目标，本发明范例可以预测和推荐博客，这样，用户可以尽早地在博客中与想法相近的 Web 参与者互动和交流，从而增加其产品或网站的知名度。在该范例中，本发明可以凸显 Web 讨论，而不是那些基于纯粹的关键词索引所找到的其他类型的相关内容，因为那些相关内容对于实现积极的线上互动这个目标来说并不是相关的。在本发明的一个典型实施方案中，这种捆绑关联的信息可以通过 API 获得。

[0031] 在一个典型实施方案中，本发明可以包括允许用户将其捆绑关联的活动和事件流个性化定制并进行访问的 Web 或移动应用。例如，该应用可以基于 Web 上的用户活动来为用户提供主题的社交图谱。本发明实施方案可允许用户查看智能捆绑起来的信息流或者

全部的未捆绑但索引过的信息流。该应用可以提供其他一些信息，例如热门概念或概念中的热门概念。在一个典型实施方案中，该应用还允许用户登录以便获取基于他们的私有数据库和账号过滤所得的内容。这些数据库和账号包括但不限于其现有的社交网络、邮箱帐号以及机构内部的数据库。在一个典型实施方案中，本发明可以将其 Web 活动索引和概念代码创建方法应用到用户的私有或公开数据，使得用户可以以统一视图查看公共和私有信息。此外，本发明的实施方案可以允许用户仅查看其私有信息。最后，本发明的实施方案可以允许用户与其他用户出于合作目的分享其活动流，包括公开或私有内容。例如，两名企业主可以分享同一个过滤后的包括其公开和私有数据的 Web 内容信息流，这样他们可以通过一个统一的视图和应用来讨论过滤后的内容。

[0032] 在一个典型实施方案中，本发明可以提供软件实施方式、云实施方式、或者能够让企业自己运行维护的软硬一体机 (appliance)，一体机既可以为了安全部署在企业的防火墙后，亦可以部署在云计算环境中。例如，组织机构可以在安全的环境下将 Web 活动索引技术应用到其自己的内部数据上。这种实施方式还可以使组织机构及其内的用户能够创建专有的概念代码 (Ticker) 或架构 (Schema)，(包括已有的和新的概念代码或架构)。这些标签和架构既可以仅被组织机构 (包括其客户和供货商) 自己使用，或者可以被公开使用。此外，本发明还可以实现闭合的反馈回路，其中的索引算法可以专门为组织机构的用户群优化。

[0033] 在一个典型实施方案中，本发明可以包含一个事件路由 (routing) 子系统，用于以可扩展的方式来分送 Web 事件。例如，路由子系统可以利用一个发布和订阅框架来以可扩展的方式将 Web 事件发送给订阅者。本发明实施方案可以支持多种协议，包括但不限于专有协议、XMPP 协议、AMQP 协议、Pubsubhub (PSHB) 协议和 RSS 云协议。通过使用一个非发布和订阅协议，或者轮询 (Polling) 协议，数据还可以经由 HTTP 请求来获得。本发明的实施方案可以针对其支持的每个协议支持对应的 API。

[0034] 在一个典型实施方案中，本发明可以支持通配符以允许程序员访问新概念或特定概念中的新概念。

[0035] 在一个典型实施方案中，本发明可以包含一个基于规则的过滤子系统，用以支持事件路由。例如，用户可以定义具体的规则说明什么时候数据应该发送过来。这样的规则例子包括但不限于：Web 活动水平或针对特定概念的 Web 活动水平，热门 Web 活动水平或针对特定概念的热门 Web 活动水平，用户参与程度或针对特定主题的用户参与程度，概念中出现的特定关键字，某网站上或者某作者生成的内容，与发现有关的任何事项、以及基于本发明的捆绑关联技术的任何情报。本发明还可以包含基于规则的优化技术，用于将数据推送给大量订阅者并且进行优化来支持大量的规则。

[0036] 本发明的一个实施方案可以支持基于信息的隐式路由，包括但不限于：用户的社交图谱，用户的资料（例如维基百科中关于用户或组织的公开信息），任何用户、组织或其网络所产生的 Web 活动。

[0037] 本发明实施方案可以包括一个应用商店。在这个应用商店中，开发者通过利用本发明所提供的数据或其拥有的任何私有数据来开发应用程序。他们可以销售和授权应用程序，或者通过这些应用程序来挣取广告收入。

附图说明

- [0038] 图 1 是根据本发明的一个典型实施方案所绘制的流程图；
- [0039] 图 2 是根据本发明的一个典型实施方案所绘制的流程图；
- [0040] 图 3 是本发明的一个实施方案中提到的不同类型捆绑关联活动的列表例子；
- [0041] 图 4 展示了本发明一个实施方案中的典型数据模型，以及
- [0042] 图 5 是根据本发明一个典型实施方案所绘制的流程图。

具体实施方式

[0043] 虽然下面的详细描述包括了很多用于说明解释目的的细节，但在本发明的范围内可以对下面的细节进行很多变化和修改。下面给出的本发明的典型实施方案不失一般性，且不会给本发明申明的权利带来任何限制。

[0044] 在过去的几年中，Web 活动、用户行为、API、API 调用、和数据的数量大幅增加。管理和采编这些大量的信息对个人和企业是一个很大的挑战。图 1 是本发明的一个典型实施方案的流程图。如图 1 所示，在一个事件驱动的架构中，Web 活动可以转变为可管理的事件（Web 事件）。在事件驱动架构中实现这种转变的重要性在于 Web 在转变为更加实时和动态的生态系统，（这与股票市场的发展如出一辙），并且需要及时采编并确定信息的相关性。

[0045] 如图 1 所示，在步骤 110 中，Web 活动可以被解析。Web 活动可以通过例如推送信息、API 或爬虫抓取的方式引入。在步骤 120 中，可以对 Web 活动进行索引成为（新的或现有的）概念。如果一个概念被识别为新的，新的概念将被创建。Web 活动可以被索引成为一个专有的数据模型，例如图 4 中所示的典型的数据模型。在步骤 130 中，可以使用一个进程从该具体的 Web 活动中识别 Web 事件。Web 事件可以具体关联这个新的 Web 事件，但也可以关联过去的和未来的 Web 活动，以及从本发明中获得的相互关联关系。

[0046] 在步骤 140 中，把历史的和其他最近的 Web 活动放到一起分析后，Web 活动和 Web 事件可以智能地捆绑在一起，并且可以相互关联，以创建一种智能的和专用的 Web 活动流。该活动流能够使用户轻松捕捉关于内容、人、以及他们感兴趣的主题的活动和相互关系。在一个典型实施方案中，可以看到对于人和内容的推荐、新的相关概念的建议、发现、以及预测。

[0047] 图 2 是根据本发明的一个示例性实施例的流程图。如图 2 所示，在步骤 210 中，Web 活动（例如来自用户“Web 参与者 Z”的评论）可以被抓取和解析。在步骤 220 中，对这个 Web 活动进行分析从中提取出一个概念“概念 Y”。这个 Web 活动，（在这种情况下是一条评论），可以索引到该概念并被存储在数据模型中（例如图 4）以捕捉所有的信息和关系。在步骤 230 中，可以从这个 Web 活动中识别出 Web 事件。在网站上评论的实例中，Web 事件可以为，例如：

- [0048] – 概念 Y 中 Web 活动的类型（即，评论）；
- [0049] – Web 参与者 Z 参与了概念 Y；
- [0050] – Web 参与者 Z 是概念 Y 中的一个新的参与者；
- [0051] – 概念 Y 中的评论的时间戳；
- [0052] – 概念 Y 中的评论的正面情绪；
- [0053] – 网页 X 活动和评论增长趋势向上；以及

[0054] -Web 参与者 Z、网页 X、情绪及概念 Y 等之间的相互关系等。
[0055] Web 活动可能涉及到在 Web 上的发生的多个典型事件，这些事件可以被存储、监视、以及和其他 Web 事件进行比较分析。

[0056] 在步骤 240 中，可以对 Web 活动和 Web 事件进行分析和捆绑关联以形成 Web 的精选集锦 (Highlight Reel) 或精典荟萃 (cliff notes) 类型的视图。上述捆绑关联可以针对感兴趣的主题。在图 2 所示的典型实施方案中，可以创建四个捆绑关联在一起的事件，这样用户可以知道 Web 活动（即，评论）是怎样与其他对时间敏感的活动关联，并获知在用户感兴趣的领域正在发生的情况的内情。

[0057] 图 3 是根据本发明的实施方案生成的捆绑关联的活动和事件的类型的一个典型表单。基于搜索引擎和社交网络的当前实现原理、或者这两者的结合，很难获得捆绑关联的活动和事件。通过突出人、内容、概念、活动水平、记录属性、和派生属性之间的独特关系，用户可以在 Web 的信息海洋中获得一个独特的、非常有吸引力的、和有价值的视角。应当说明的是，这仅仅是一个演示如何应用 Web 事件和索引的 Web 活动的示例。

[0058] 典型推荐事件 310 包括：

- [0059] - 推荐：基于您的 [Facebook] 账户的隐含兴趣，建议您关注概念代码 XYZ；
 - [0060] - 推荐：基于您的 [Twitter] 账户的粉丝，建议您关注用户 Z；
 - [0061] - 推荐：基于 [Facebook] 上您的朋友的讨论和活动，建议您研读 [网页链接 URL]；
 - [0062] - 推荐：XYZ 博客 /URL 显示了很多关于该概念代码的早期活动，加入评论对市场营销会有所帮助；以及
- [0063] - 推荐：相关的概念代码 123 显示了比平时更高的参与度，参与讨论对市场营销应该有价值。

[0064] 典型有影响力的用户事件 320 包括：

- [0065] - 有影响力的用户：用户 A 在该概念代码中变得日益活跃；以及
- [0066] - 有影响力的用户：以下有影响力的用户正在对 [标签] 发表推文（微博）。

[0067] 典型位置事件 330 包括：

- [0068] - 位置：纽约市出现了很多关于该概念代码的活动；
- [0069] - 位置：当前有很多有影响力的用户聚集在纽约的 ABC 咖啡店；以及
- [0070] - 位置：当前有大量关于纽约市的 JFK 机场的文章涌现。

[0071] 典型预测事件 340 包括：

- [0072] - 预测：在该主题中，用户 A 将变成一个有影响力的用户；
- [0073] - 预测：由于该概念代码的关键有影响力的用户的参与，XYZ 博客将出现大量的流量；以及

[0074] - 预测：基于早期出现的异常活动，相关概念代码 ABC 有望成为一个顶级热门的概念代码。

[0075] 典型发现事件 350 包括：

- [0076] - 发现：一个与您的兴趣有关的新概念 / 概念代码已经出现；
- [0077] - 发现：发现了一个有影响力的用户们正在积极参与的新博客；以及
- [0078] - 发现：在相关概念代码 XYZ 中，情绪（态度）有了突然而且明显的变化，这个现象值得进一步关注。

[0079] 典型讨论事件 360 包括：

[0080] - 讨论：有大量和本概念代码相关的关于〔关键字标签〕的讨论出现；

[0081] - 讨论：用户 D 参与许多关于该概念代码的活动。查看推文（链接）；以及

[0082] - 讨论：在您的社交网络中的两个人（用户 A 和用户 B）正在进行与该概念代码有关的讨论。

[0083] 典型活动事件 370 包括：

[0084] - 活动水平：有大量与网站 X 有关的推荐（Diggs）出现；

[0085] - 活动水平：有大量与网站 Y 有关的推文出现；以及

[0086] - 活动水平：该概念代码中出现了很多一般不参与这个主题的用户的活动，这显示了该主题更广泛的吸引力。

[0087] 图 4 展示了基于本发明实施方案的一个典型数据模型。如图 4 所示，这个数据模型可以捕捉并能够实现如下这些个体之间的相互关系的映射：关键字 410、概念 420、概念的属性 425、Web 参与者 430、Web 参与者的属性 435、数据记录 440（例如 URL、推文、微博、消息、聊天、评论、API 或 API 调用、电子邮件、数据文件、电话、音频、视频，或者未来可获得的任何类型的数据记录、数据记录的属性）以及派生属性 450（例如内部监视的 Web 事件）。这种独特的关系图谱支持独特的分析，尤其是在事件驱动架构内进行处理的时候。

[0088] 图 5 是一个基于本发明的典型实施方案的流程图。如图 5 所示，Web 活动可以来源于推送信息处理，信息抓取，API 或其他方法，并通过负责实时信息抓取、推送信息处理、和解析的模块 505（“抓取组件”）来进行处理。Web 活动可以被解析并传递到概念索引子系统 510，亦可以被传递到社交图谱分析子系统 525，如下所述。作为一个可选项，抓取组件 505 可以包含一个监视组件（图中未显示），以监视内容的更新。抓取组件 505 可以按特定频率、或在特定时间、或在特定事件发生时安排抓取活动。概念索引子系统 510 可以通过应用语义分析、聚类分析和模糊匹配技术来提取主题以索引 Web 活动。这些主题可遵循一种“自组织”方式以反映线上内容的变化，与之形成对照的是通常使用的一种自上而下的分类架构方式。本发明支持使用这两种方式的任意一种。概念的例子可以是“猪流感”、“实时搜索”、“巴拉克·奥巴马”以及“微软收购雅虎”。主题中的字数没有限制。

[0089] 语义分析模块 511 可以用来进一步分析 Web 活动，该语义分析模块把同义词和多义词等因素整合分析。与关键字不同，这样处理的好处在于，允许概念捕捉多种含义，从而更好地反映其对应的 Web 活动。作为一个类比，如果一个代表微软的股票代码不考虑与 Microsoft、MSFT、Microsoft Corporation、Micro-soft 等相关的消息，那么对用户而言监视这个股票代码的意义就大大减小了，因为这会丢失大量有用的信息。

[0090] 情绪分析子系统 512 可以用来分析 Web 活动里的情绪是正面的、负面的还是中性的。这可以独立地或与索引到概念中的其他 Web 活动情绪信息一起为本发明提供有价值事件信息。作为一个可选项，分类子系统 513 可以用来进一步分析 Web 活动。分类子系统 513 可以分析 Web 活动的权威性以确定其是否是垃圾信息、非常权威，或者介于两者之间。分类子系统 513 还可以基于不同的分类架构体系对 Web 活动的内容进行分类。这些分类架构体系包括但不限于：体育、政治、娱乐、游戏和健康等，或者新闻、博客、微博、图像、视频和音频等，或者英语、西班牙语、汉语和法语等语言分类，或者新颖内容与陈旧信息等，或者色情与非色情等，或者购买意向等。

[0091] Web 活动可以通过分类子系统 513 推回到概念索引子系统 510，并且可选择被推进到影响力排序子系统 535 以计算 Web 活动的影响力。影响力排序子系统 535 可以将概念索引子系统 510 中识别出的概念和 Web 活动与社交图谱分析子系统 525 中的分析结果结合。社交图谱分析子系统 525 可以识别 Web 活动中的 Web 参与者，并且可以分析隐性和显性的社交图谱关系。例如，该社交图谱分析子系统 525 可以基于博客中互相评论的 Web 参与者、社交网络中的显性关系和信息交流、以及社交网络中的关系变化来确定隐性关系。

[0092] 社交图谱分析子系统 525 可以将信息传送到概念索引子系统 510 和影响力排序子系统 535。影响力排序子系统 535 可以为每个概念构建社交图谱。针对一个概念，影响力排序子系统 535 可以识别哪些参与者是积极参与或者轻度参与。该影响力排序子系统 535 可以监视概念内 Web 参与者的活动水平随时间推移的变化，从而识别哪些 Web 参与者的影响力在增强而哪些 Web 参与者的影响力在削弱。该影响力排序子系统 535 可以追踪概念内 Web 参与者之间信息流动的路径，以及信息被传送的方法（如评论、推文等），同时考虑特定概念和内容传播所需的时间。

[0093] 当内容在 Web 参与者之间传送时，可以应用到一种独特的排名计分方法。这个分数可以同时被应用到 Web 参与者和内容本身。例如，如果内容在有影响力的人之间被快速地传送，那么这个内容可以获得非常高的分数并且对外部的 Web 参与者而言将很有可能是非常相关和重要的。在这种情况下，本发明实施方案可以通知 Web 参与者相关信息的存在。如果有影响力的人将内容传送给影响力小的人，影响力小的人的影响力就会上升，因其现在更可能拥有具有影响力的信息。最后，信息路径可以被存储并用于衡量相关性，这样如果未来出现相似的路径，那么这个信息是相关的概率就比较高。这种相关性测定方法是用于预测天气、风暴和飓风的常用技术。对历史数据进行概率分析可以帮助预报和预测未来发生的事件。

[0094] Web 活动索引子系统 515 可以将概念索引子系统 510 和影响力排序子系统 535 中的数据结合，并将这些数据规范化 (normalized) 存入数据仓库 520。该数据仓库 520 可以支持例如图 4 所示的数据模型。

[0095] 在 Web 活动索引子系统 515 对 Web 活动进行索引的同时，可以将 Web 活动从概念索引子系统 510 传送到概念代码管理子系统 530。概念代码管理子系统 530 可以创建概念代码（相当于标签或可编程主题标签）以反映概念。如果新概念被识别出来，概念代码管理子系统 530 可以创建新的概念代码以反映这一概念。概念代码管理子系统 530 可以将推荐的概念代码推送给用户以提供用于发现的强大工具。例如，如果有新的相关概念代码与用户正在关注的概念高度相关，该代码管理子系统 530 可以建议用户也关注新的代码。该概念代码可以被传送到概念代码补充和丰富子系统 531 对其进行信息补充和丰富。

[0096] 概念代码补充和丰富子系统 531 可以使用专有的知识库和第三方数据源，包括但不限于：人工采编而成的数据来源（例如维基百科和 Freebase）、结构化的数据来源（例如 Wolfram），以及用户定义的元数据，其中用户可以创建私有和公共内容级别和类别。这为用户订阅提供了更好的概念代码内容分类。例如，冠蓝鸦 (bluejay) 可以是一种鸟，也可以是一只运动队的名字。使用信息补充和丰富的方法，本发明能够分离（有歧义的）内容，使得每种内容有不同的类别。还有一种用户自定义的情况，其中用户可以提供关键字标签和“Web 活动标签”以指示本发明实施方案怎样索引 Web 活动。用户自定义的元数据可以在私

有环境中使用,例如在企业内,或者对外公开使用。应当注意的是,在某些情况下,通过信息丰富处理,概念代码可以等同为一个数值。例如,概念代码可以代表一个城市的人口,这相当于一个数字。

[0097] 经过概念代码补充和丰富子系统 531 处理后的数据可以被回传到概念代码管理子系统 530,然后可以被存储在数据仓库 520 中,可以被推送到 API590、被推送到 Web 流管理子系统 560、被推送到配置和管理子系统 555,和 / 或被推送到 Web 活动及事件描述生成子系统 575。需要注意的是,在每种情况下,用于表示数据流的线条是双向的,以反映用户定义的数据和概念代码的订阅情况。

[0098] 一旦数据被存储在数据仓库 520 中且支持用户订阅的概念代码已经被创建,基于用户的需求和数据类型,存在很多使用案例。这些使用案例中的一个或者全部都可以通过本发明实施方案来实现。

[0099] 在一个典型实施方案中,通过概念代码索引到相关概念的全部 Web 活动流可以经由信息流管理子系统 560 推送给用户或者商户。流管理子系统 560 可以管理流订阅和用户的过滤规则,并将数据推送到 API 590。在其他实施方案中,开发者可以经由配置和管理子系统 555 来订阅数据流。配置和管理子系统 555 可以包括图形用户界面和基于规则的过滤子系统 550 用以基于规则来过滤 Web 活动。

[0100] 本发明的典型实施方案可以将数据仓库 520 中的数据传送到 Web 事件创建子系统 565。该 Web 事件创建子系统 565 可以将基本的 Web 活动转换为可以被监视的独特事件。Web 事件可以 :i) 被存储在数据仓库 520 中 ;ii) 被传送到 Web 活动和事件排序子系统 540,在那里 Web 活动和事件被排序,然后被回传到 Web 事件创建子系统 565 ;或 iii) 通过 Web 事件捆绑关联子系统 570 被捆绑关联和分析,然后通过 Web 活动和事件描述生成子系统 575 生成描述。Web 事件捆绑关联子系统 570 以及 Web 活动和事件描述生成子系统 575 可以生成图 3 中列出的典型捆绑关联的活动和事件。Web 活动和事件描述生成子系统 575 可以将捆绑关联的事件推送到 API 590。这是一种双向流,用以反映用户反馈和请求。

[0101] 在一个典型实施方案中,通过 Web 事件创建子系统 565 创建并存储在数据仓库 520 中的 Web 事件可以被传送到复杂事件处理和分析子系统 580 (“CEP”)。由于本发明的实施方案可以将基本的 Web 活动转换为 Web 事件,因此可以应用事件驱动分析技术来分析事件。该子系统可以使用面向计算 (Computation-Oriented) 的 CEP 和面向检测 (Detection-Oriented) 的 CEP 这两种技术。CEP 子系统 580 可以使用以下技术例如事件关联和抽象、多事件等级的复杂模式的检测、以及事件之间的关系,例如因果关系、会员资格、时机巧合 (timing) 和事件驱动进程。CEP 子系统 580 可以推断和预测关系、事件、相关性以及未来 Web 活动。

[0102] 传统的搜索引擎通过衡量网页的人气来衡量大众的智慧,通过创建和分析事件,本发明实施方案可以领先于大众来预测智慧。以股市为例进行类比,股市上股票的价格反映了大众的群体智慧 (这是效率市场理论的功能),但算法交易技术利用事件的模式和相互关联性以预测股票市场中的高概率股票价格走向趋势。通过将 Web 活动转换为一个用可以被监视和分析的事件描述的框架,本发明可以将 Web 从基于内容的模型转化为基于可量化事件的模型。

[0103] CEP 子系统 580 可以将数据推送到 API 590,推回到数据仓库 520,或者推送到 Web

事件创建子系统 565, 那里可以处理新的 CEP 事件。

[0104] 在 API 590 中, 数据可以被直接访问或者被推送到开发者框架 591、Web 应用 592、移动应用 593、事件路由 (event-routing) 分配框架 594 中, 或者被推入到云上的软硬一体机或服务实例 595。软硬一体机 595 使得企业或商户可以使用本发明中所描述的任何组件供其自身使用和定制数据。

[0105] Web 应用 592 和移动应用 593 的示例包括但不限于 :Web 活动流, 它提供了 Web 上围绕感兴趣的概念的精华集锦 ; 目录应用, 可以用来展示 Web 参与者、概念、内容和数据记录 (URL) 之间的关系以及这些关系随着时间的变化。

[0106] 显然, 在不脱离本发明的精神和范围的前提下, 专业人士可以对 “在事件驱动架构下对 Web 活动索引、排序和分析的系统和方法”的本发明实施方案进行各种修改和变型。因此, 在所述修改和变型落入本发明的权利要求及其等同体所限定的保护范围的情况下, 本发明实施方案旨在涵盖上述针对本发明所做的所有修改和变型。

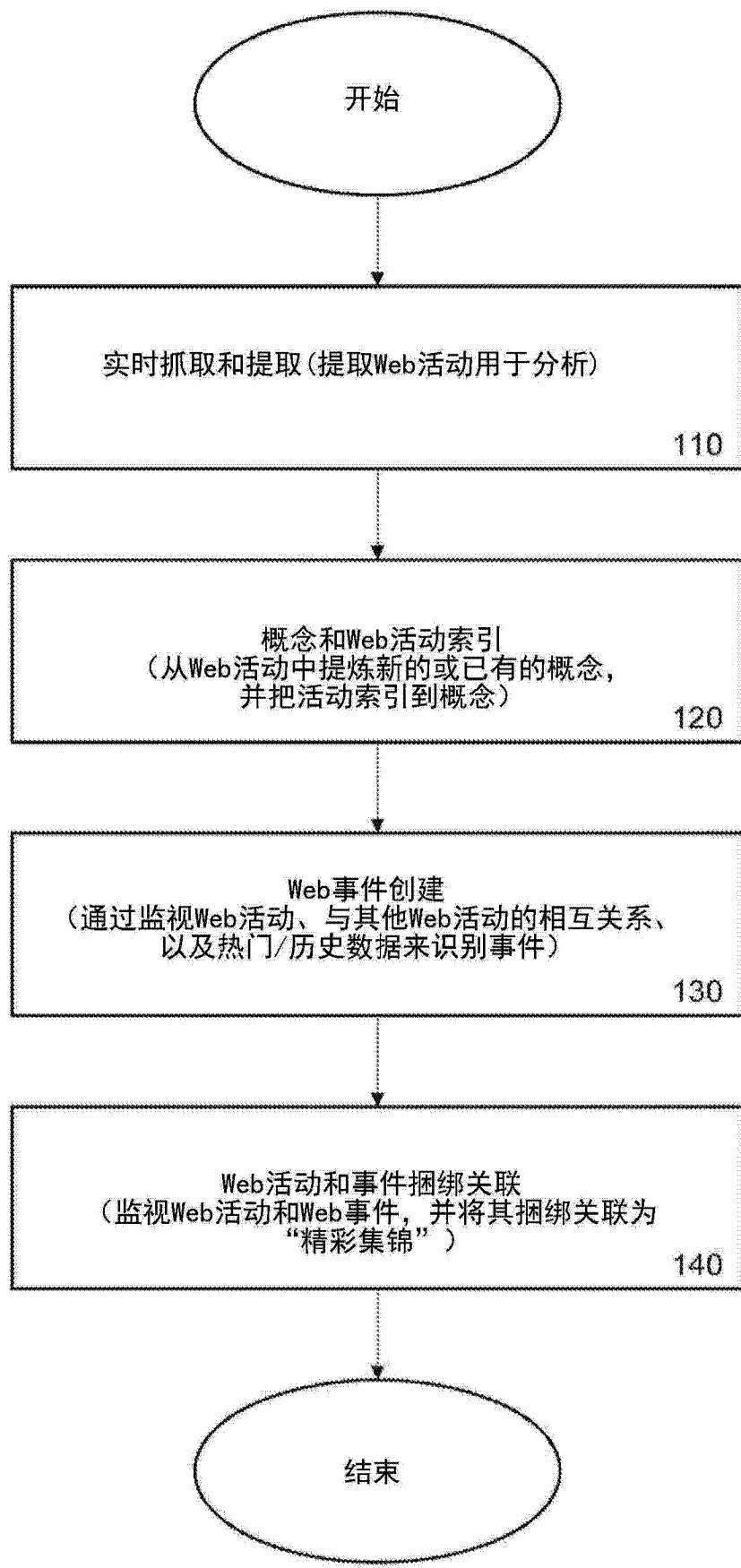


图 1

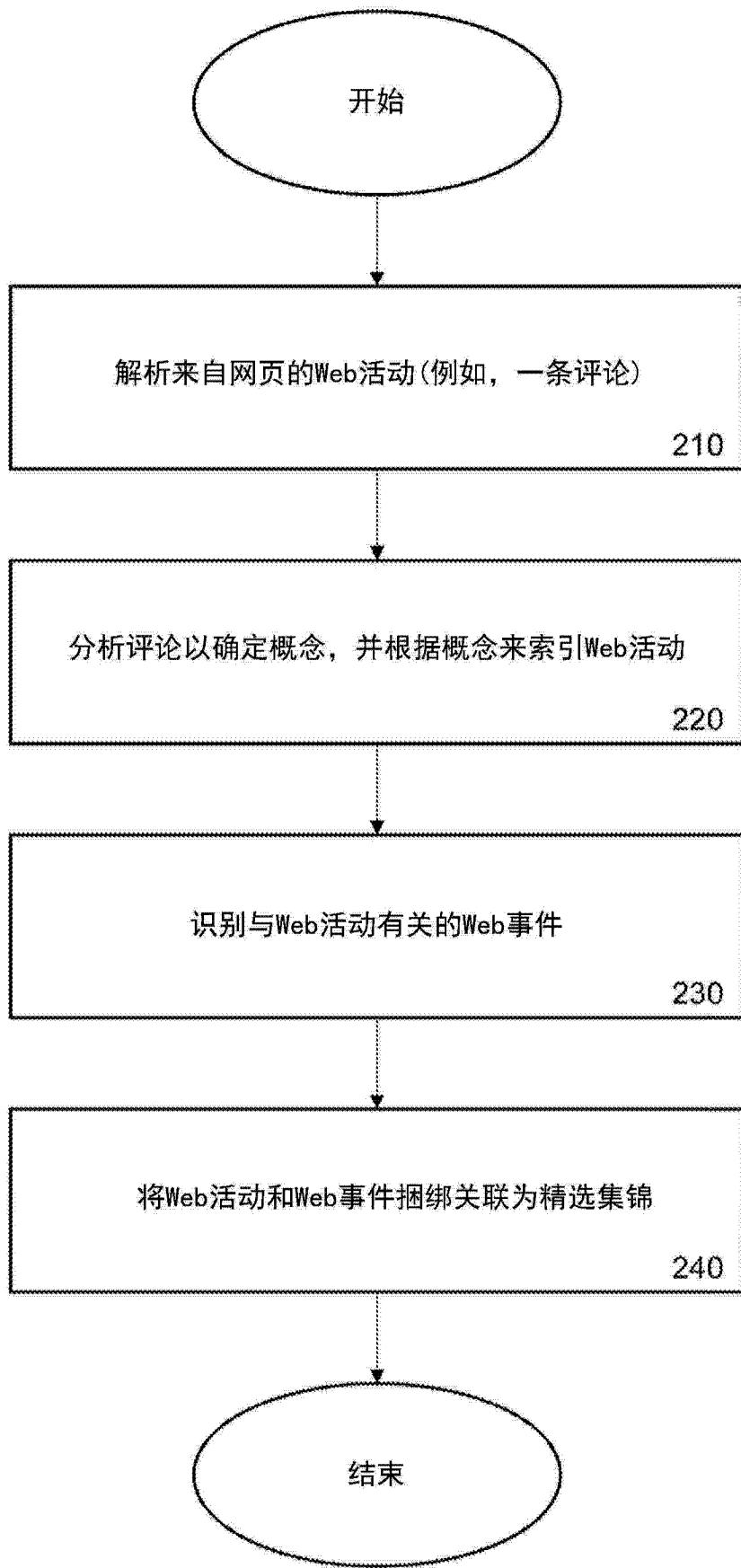


图 2

推荐事件示例（310）：

- 推荐：基于您的[Facebook]账户的隐含兴趣，建议您关注概念代码XYZ；
- 推荐：基于您的[Twitter]账户的粉丝，建议您关注用户Z；
- 推荐：基于[Facebook]上您的朋友的讨论和活动，建议您研读[网页链接URL]；
- 推荐：XYZ博客/URL显示了很多关于该概念代码的早期活动，加入评论对市场营销会有所帮助；以及
- 推荐：相关的概念代码123显示了比平时更高的参与度，参与讨论对市场营销应该有价值。

有影响力的用户事件示例（320）：

- 有影响力的用户：用户A在该概念代码中变得日益活跃；以及
- 有影响力的用户：以下有影响力的用户正在对[标签]发表推文（微博）。

位置事件示例（330）：

- 位置：纽约市出现了很多关于该概念代码的活动；
- 位置：当前有很多有影响力的用户聚集在纽约的ABC咖啡店；以及
- 位置：当前有大量关于纽约市的JFK机场的文章涌现。

预测事件示例（340）：

- 预测：在该主题中，用户A将变成一个有影响力的用户；
- 预测：由于该概念代码的关键有影响力用户的参与，XYZ博客将出现大量的流量；以及
- 预测：基于早期出现的异常活动，相关概念代码ABC有望成为一个顶级热门的概念代码。

发现事件示例（350）：

- 发现：一个与您的兴趣有关的新概念/概念代码已经出现；
- 发现：发现了一个有影响力的用户们正在积极参与的新博客；以及
- 发现：在相关概念代码XYZ中，情绪（态度）有了突然而且明显的变化，这个现象值得进一步关注。

讨论事件示例（360）：

- 讨论：有大量的与本概念代码相关的关于[关键字标签]的讨论出现；
- 讨论：用户D参与许多关于该概念代码的活动。查看推文（链接）；以及
- 讨论：在您的社交网络中的两个人（用户A和用户B）正在进行与该概念代码有关的讨论。

活动事件示例（370）：

- 活动水平：有大量的与网站X有关的推荐（Diggs）出现；
- 活动水平：有大量的与网站Y有关的推文出现；以及
- 活动水平：该概念代码中出现了很多一般不参与这个主题的用户的活动，这显示了该主题更广泛的吸引力。

图 3

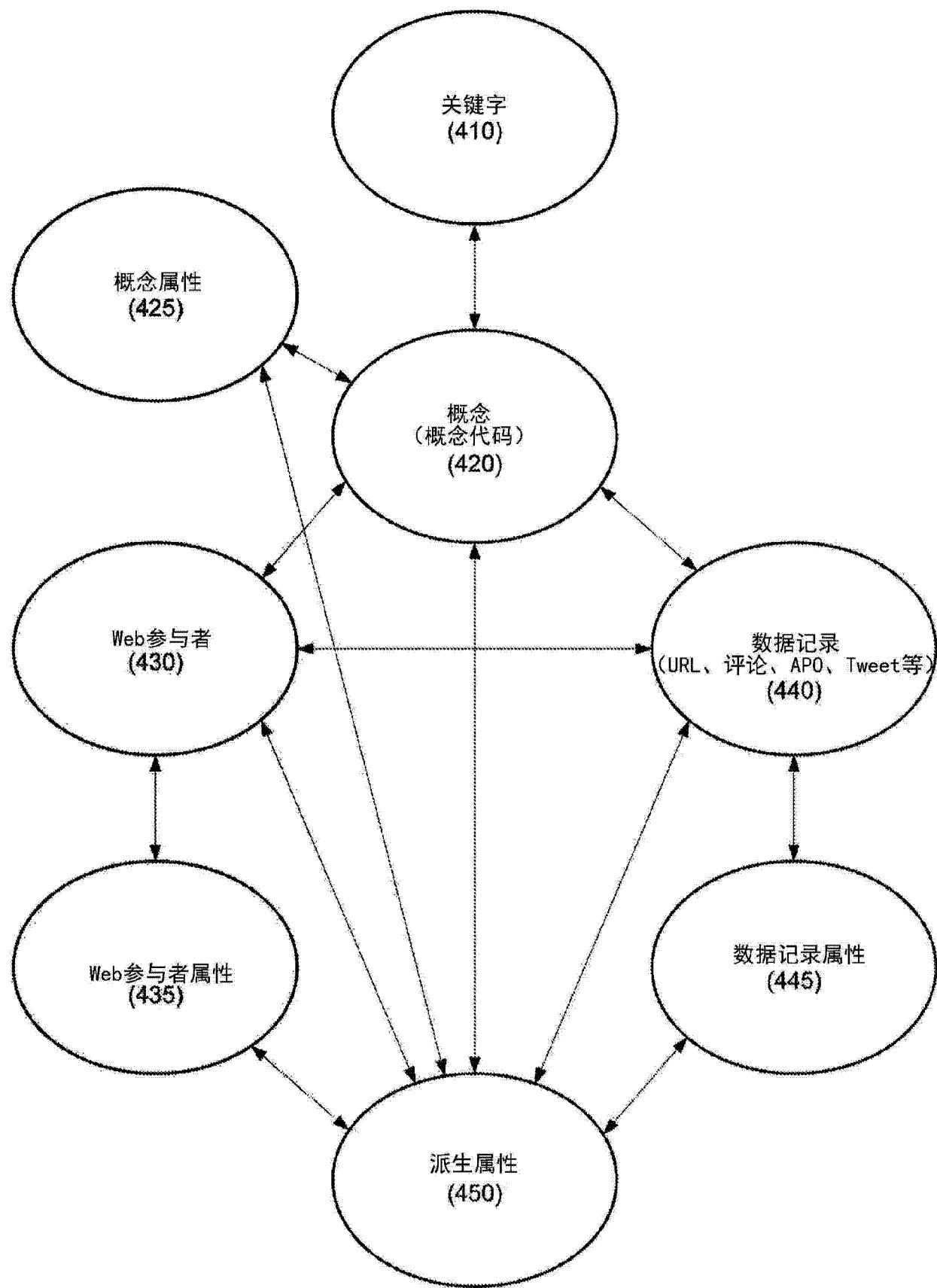


图 4

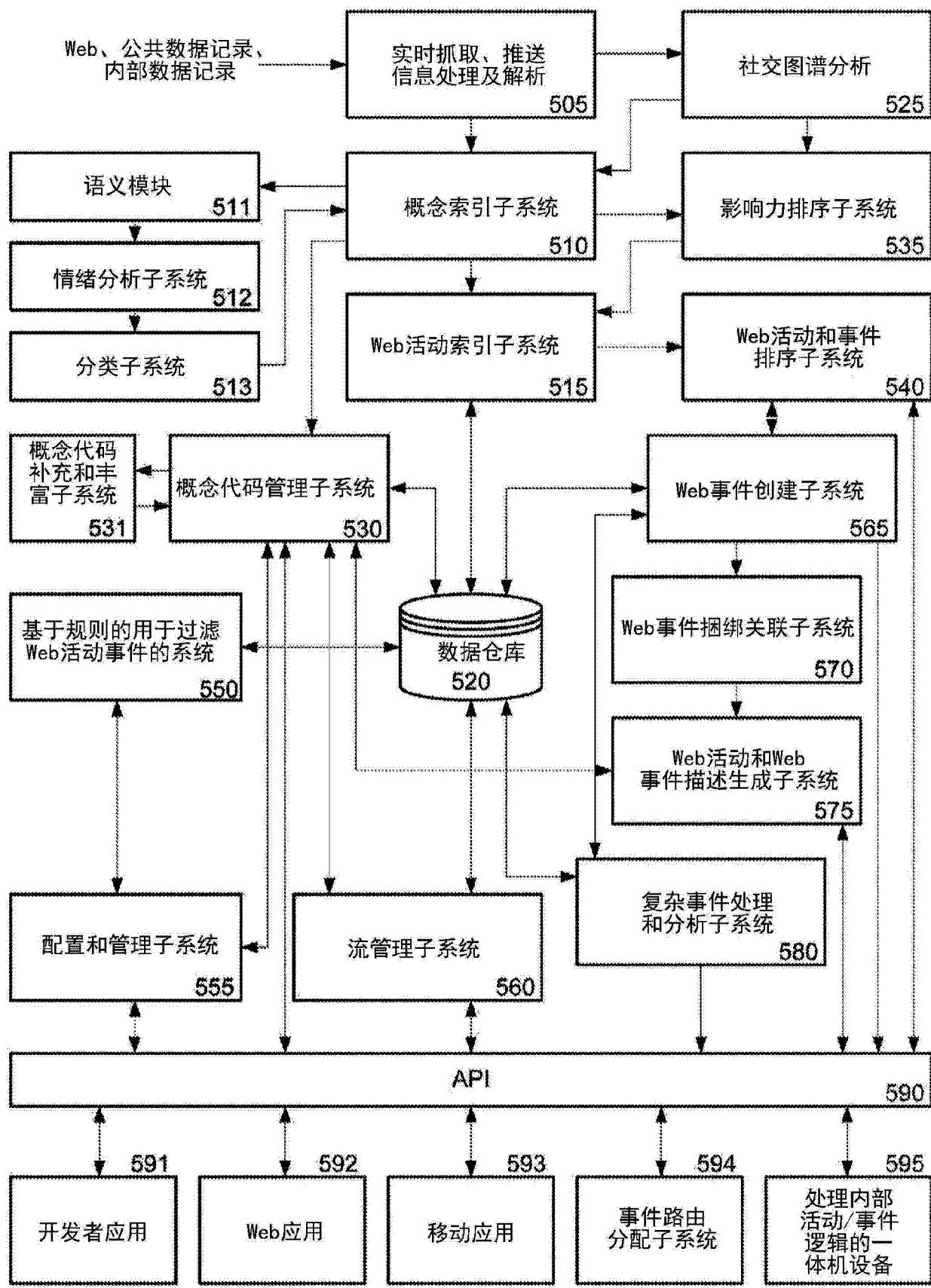


图 5