



(12) 发明专利申请

(10) 申请公布号 CN 101779180 A

(43) 申请公布日 2010.07.14

(21) 申请号 200880102339.5

(51) Int. Cl.

(22) 申请日 2008.08.08

G06F 3/00(2006.01)

G06F 13/00(2006.01)

(30) 优先权数据

60/954,677 2007.08.08 US

12/188,850 2008.08.08 US

(85) PCT申请进入国家阶段日

2010.02.08

(86) PCT申请的申请数据

PCT/US2008/072660 2008.08.08

(87) PCT申请的公布数据

W02009/021198 EN 2009.02.12

(71) 申请人 贝诺特公司

地址 美国加利福尼亚州

(72) 发明人 斯科特·布雷夫 杰克·贾

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 颜涛 郑霞

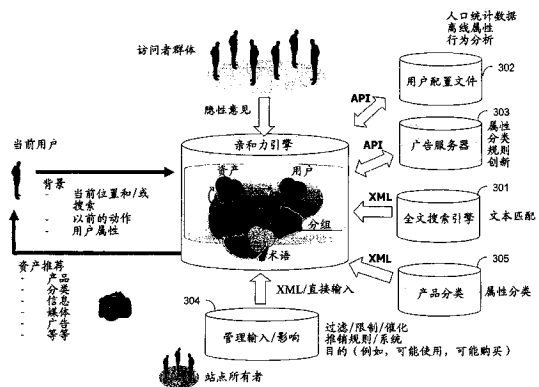
权利要求书 4 页 说明书 22 页 附图 30 页  
按照条约第19条的修改 4 页

(54) 发明名称

基于背景的内容推荐的方法和设备

(57) 摘要

从企业内部和周边的人们开始,储存在人们的大脑中的在他们日常的行为中所展示的专业知识和工作模式被检测和捕获。一个基于行为的知识索引因而被创建,其用来产生专家指导的、个性化的信息。



1. 一种用于内容推荐的由计算机实施的方法,包括以下步骤:  
所述计算机接收用户动作所导致的信息,并识别当前用户的背景;  
基于所述识别,所述计算机找到适当的对等组,所述对等组代表在所述识别的背景中与所述当前用户最像的其他用户;  
所述计算机找到所述对等组识别为与所述当前背景最相关的内容;以及  
所述计算机产生输出,所述输出包括用于显示给所述当前用户的推荐的内容。
2. 如权利要求 1 所述的方法,还包括以下步骤:  
所述计算机通过以下操作产生全面的行为指纹:  
考虑包括多种多样的用户行为的输入;  
将所有的这些用户行为与对应于所述当前用户在其他页面上的行为的储存信息进行互相关,并与群体中其余部分和识别出的对等体在当前页面上的行为进行互相关;  
根据此分析,所述计算机产生所述当前用户在一段特定的内容中已经找到价值的可能性;以及  
所述计算机储存所述可能性供在学习系统中使用。
3. 一种由计算机实施的以背景为中心的内容推荐方法,其包括以下步骤:  
所述计算机基于代表网站上的用户动作的输入,为所述网站上的用户建立当前背景;  
所述计算机用嵌入所述网站中的观察者标记捕获所述用户的当前背景信息;  
所述计算机基于由所述观察者标记所收集的所有的用户行为确定所述用户的兴趣;  
所述计算机使用全面行为指纹技术来分析所述行为;  
所述计算机将所有的结果信息作为所述用户的当前背景矢量储存在存储器中,所述背景矢量包括在每一个条目上都带有权重的、术语和文档的混合矢量,所述权重反映该术语或文档反映所述用户的当前背景的强烈程度;  
当用户输入搜索术语和 / 或点击导航链接时,所述计算机增加对应于被输入或被点击的术语和短语的所述背景矢量的条目,以捕获所表达的兴趣;  
当用户动作进一步移动成为过去时,所述计算机减少或衰减相应的条目;  
所述计算机增加用户点击的或基于用户的隐性动作所确定的用户表示感兴趣的相应的矢量条目文档;以及  
根据前述操作,所述计算机产生所述用户的当前背景的代表来作为背景矢量,用于对所述用户进行内容推荐。
4. 如权利要求 3 所述的方法,还包括以下步骤:  
基于从聚合智慧所学习到的亲和力和联系,所述计算机将所述背景矢量扩展和改进为意图矢量,所述聚合智慧从对群体作为一个整体的长期的观察中收集;  
其中所述意图矢量通过查看在所述背景矢量中的术语和文档与其他术语的亲和力来创建。
5. 如权利要求 4 所述的方法,还包括以下步骤:  
所述计算机利用亲和力引擎确定亲和力,所述亲和力允许将所述背景矢量扩展为意图矢量;  
其中所述亲和力引擎通过观看用户群体在网站上的所有的隐性行为并通过应用行为指纹和使用等级算法来确定兴趣和联系,从而学习文档和术语、文档和文档、术语和术语、

以及用户到其他用户、文档、和术语之间的连接。

6. 如权利要求 5 所述的方法,还包括以下步骤:

所述计算机识别共享对当前意图的亲活力的用户的对等组、以及在该意图的背景中展示出与所述当前用户最像的行为的那些用户,所述对等组由用户矢量代表,其中在所述用户矢量中的每一个用户条目可具有权重,该权重指示他在当前背景中对于所述当前用户是对等体的强烈程度。

7. 如权利要求 6 所述的方法,还包括以下步骤:

所述计算机识别哪些文档在当前意图中对所识别的对等体具有最高的亲和力;

其中所述亲和力引擎查看与那些对等体相关的、根据他们的对等体权重被加权的活动性矢量、术语 - 文档矩阵以及下一步矩阵,并聚合地使用这些来计算对所述当前意图具有最高的亲和力的那些文档;以及

其中所述文档是未过滤的推荐。

8. 如权利要求 7 所述的方法,还包括以下步骤:

所述计算机向所述亲和力引擎请求关于所述推荐中的每一个推荐的群体信息;

其中所述信息与其他资产信息结合并被显示给用户以帮助其了解推荐所隐含的群体智慧。

9. 如权利要求 8 所述的方法,其中所显示的信息包括以下中的任意项:在文档中找到价值的用户总数、将文档与当前背景和 / 或意图关联起来的对等体的数量、以及对等群体已经将之与一文档相关联的术语和 / 或短语。

10. 如权利要求 9 所述的方法,其中,所述术语和 / 或短语包括代表所述群体已经关联到一文档的术语的虚拟公众分类;

其中所述虚拟公众分类由所述亲和力引擎基于所述用户群体的隐性动作自动创建。

11. 一种用于内容推荐的设备,包括:

用于接收输入信息以在识别当前用户的背景中使用的装置;

用于基于所述识别而找到适当的对等组的装置,所述对等组代表在所述识别的背景中与所述当前用户最像的其他用户;

用于找到所述对等组识别为与所述当前背景最相关的内容的装置;以及

用于产生包括推荐给所述当前用户的内容的输出的装置。

12. 如权利要求 11 所述的设备,还包括:

用于通过以下操作产生全面行为指纹的装置:

考虑包括多种多样的用户行为的输入;

将所有的这些用户行为与所述当前用户在其他页面上的行为、以及与群体中其余部分和所识别的对等体在当前页面上的行为进行互相关;

根据此分析,产生所述当前用户在一段特定的内容中已经找到价值的可能性;以及

储存所述可能性供学习系统使用。

13. 一种以背景为中心的内容推荐设备,包括:

用于接收代表用户动作的输入并用于根据所述输入在网站上建立当前用户背景的设备;

被嵌入所述网站用于捕获所述用户的当前背景信息的观察者标记;

用于基于由所述观察者标记所收集的各种的行为确定所述用户的兴趣的装置；

用于使用全面行为指纹技术来分析所述行为的装置；

用于将所有的结果信息作为所述用户的当前背景矢量储存在存储器中的装置，所述背景矢量包括在每一个条目上都带有权重的、术语和文档的混合矢量，所述权重反映该术语或文档反映所述用户的当前背景的强烈程度；

用于当用户输入搜索术语和 / 或点击导航链接时增加对应于被输入或被点击的术语和短语的所述背景矢量的条目以捕获所表达的兴趣的装置；

用于当用户动作进一步移动成为过去时减少或衰减相应的条目的装置；

用于增加用户点击的或基于用户的隐性动作所确定的用户表示感兴趣的相应的矢量条目文档的装置；以及

用于根据前述操作产生所述用户的当前背景的代表作为背景矢量用于对所述用户进行推荐的装置。

14. 如权利要求 13 所述的设备，还包括：

用于基于从聚合智慧所学习到的亲和力和联系将所述背景矢量扩展和改进为意图矢量的装置，所述聚合智慧从对群体作为一个整体的长期的观察中收集；

其中所述意图矢量通过查看在所述背景矢量中的术语和文档与其他术语的亲和力来创建。

15. 如权利要求 14 所述的设备，还包括：

用于利用亲和力引擎确定亲和力的装置，所述亲和力允许将所述背景矢量扩展为意图矢量；

其中所述亲和力引擎通过观看用户群体在网站上的所有的隐性行为并通过应用行为指纹和使用等级算法来确定兴趣和联系，从而学习文档和术语、文档和文档、术语和术语、以及用户到其他用户、文档、和术语之间的连接。

16. 如权利要求 15 所述的设备，还包括：

用于识别共享对当前意图的亲和力的用户的对等组以及在该意图的背景中展示出与所述当前用户最像的行为的那些用户的装置，所述对等组由用户矢量代表，其中在所述用户矢量中的每一个用户条目可具有权重，该权重指示他在当前背景中对于所述当前用户是对等体的强烈程度。

17. 如权利要求 16 所述的设备，还包括：

用于查看哪些文档在当前意图中对所识别的对等体具有最高的亲和力的装置；

其中所述亲和力引擎查看与那些对等体相关的、根据他们的对等体权重被加权的活动性矢量、术语 - 文档矩阵以及下一步矩阵，并聚合地使用这些来计算对所述当前意图具有最高的亲和力的那些文档；以及

其中所述文档是未过滤的推荐。

18. 如权利要求 17 所述的设备，还包括：

用于向所述亲和力引擎请求关于所述推荐中的每一个推荐的群体信息的装置；

其中所述信息与其他资产信息结合并被显示给所述用户以帮助其了解推荐所隐含的群体智慧。

19. 如权利要求 18 所述的设备，其中所显示的信息包括以下中的任意项：在文档中找

到价值的用户总数、将文档与当前背景和 / 或意图关联起来的对等体的数量、以及对等群体已经将之与一文档相关联的术语和 / 或短语。

20. 如权利要求 19 所述的设备, 其中, 所述术语和 / 或短语包括代表所述群体已经关联到一文档的术语的虚拟公众分类;

其中所述虚拟公众分类由所述亲和力引擎基于所述用户群体的隐性动作自动创建。

## 基于背景的内容推荐的方法和设备

[0001] 相关申请的交叉引用

[0002] 本申请要求 2008 年 8 月 8 日提交的美国专利申请第 12/188,850 号、以及 2007 年 8 月 8 日提交的美国临时专利申请第 60/954,677 号的权益,两者通过引用而整体并入此处。

[0003] 发明背景

### 技术领域

[0004] 本发明涉及以电子方式获取信息。更具体地,本发明涉及一种用于基于背景的内容推荐的方法和设备。

[0005] 现有技术描述

[0006] 在电子网络中伴随着找到信息的一个问题是关于人们怎样尽快和尽可能有效地与满足他们需要的信息/产品/服务连接。这是自万维网(WorldWide Web)开始以来,网页和搜索引擎的主要目标之一。未能做到如此会导致在电子商务、电子旅行以及电子营销站点的情况下失去业务,使电子支持站点上的顾客觉得沮丧,其有可能之后呼叫顾客支持,因而浪费了许多公司的金钱,使电子媒体站点上的观众/读者失去兴趣,其迅速放弃该站点,因此对于广告收入、以及内部网上的非生产性雇员丧失了机会。

[0007] 网站设计是解决信息发现问题的人工尝试:以一种设计者设想会帮助用户找到他们正在寻找的信息的方式来组织信息。这种方式虽然在某些情况下有效,但是以这种方式试图找到信息常常是缓慢和低效的,因为用户采取在站点上到处浏览以寻找他们需要的信息。实际上,如果在三次点击内,用户没有找到所寻找的信息,多数会放弃这个站点。一个问题是该站点是静态的。最近几年,作为缓解这个问题的尝试,Web 分析学已经出现。设计者可看到在他们的站点上发生的所有动作,并且将其收集在报告中,报告旨在提供怎样更有效地重新设计或重新配置站点的一些指导。这虽然提供了某些益处,但是所提供的信息常常是含糊不清的,并且只提供用于改进的提示而不是具体建议。在最好的情况下,这个进程也是乏味的,要求大量的人工精力,因为设计者要与认知一致来重新设计站点,并且需要很长的时间。因而反馈环路是缓慢而低效的。

[0008] 自动内容推荐是在 Web 存在的非常早期出现的一种完全不同的策略。搜索引擎诸如 Google、Yahoo 以及 Ask 是这种技术的常见代表。其基本思想是,用户以搜索查询的形式明确描述他们正在寻找什么,而自动进程尝试识别最匹配他们的查询的内容片段,最经常的是一个 Web 页。这种做法等于查看所有可能的文档和推荐目标查询在文本中以最高频率发生的那些地方,即关键字匹配。这项基本技术的现代改编版本添加了复杂的层,例如自然语言处理,但是这些方式的关键仍然是使用内容自身的属性,例如文档中的词,以确定最终的关联等级。这代表了内容推荐的第一个以内容为中心的阶段(见图 1)。

[0009] 存在这种方式的许多变化,最显著地,包括元标记。在这种方式中,内容创建者选择少量术语来描述该内容。这些术语被嵌入该内容中,常常作为 HTML 元标记,但是不一定对内容的顾客可见。这是允许搜索引擎搜索不是基于文本的内容的一种方式,如视频剪辑。这种方式在 20 世纪 90 年代末非常普遍,但是由于保持该元标记为最新的以及与对内容的

修改同步所需要的巨大的精力,这种方式已经不受欢迎。

[0010] 在许多方面,这个第一个以内容为中心的方法表面上很有意义,即,如果你想要推荐内容,考虑内容自身。这种方式的一个关键问题是其常常带回大量可能是相关的但是没有用处的文档。许多文档可能展示出强烈的关键字匹配,但是已过时或没有与用户当前的兴趣真正相关。如果用户在前几个结果中没有找到有用的结果,他们最有可能会放弃该搜索。

[0011] 关键字匹配并不真实地反映我们在真实世界中怎样最有效地找到信息。在每天的日常生活中,找到我们正在寻找的信息 / 产品 / 服务的最好的方式是询问知道以正确的方向指点我们的人。因而内容推荐的第二个阶段将焦点从内容转移到用户 (见图 1)。Google 的“PageRank”算法,虽然我们把它放在阶段 1,但实际上是一种关注阶段 2 的到来的过渡技术。网页等级算法的突破是不仅考虑网页自身的内容,而且考虑其由其他网站设计者从其他网页怎样被链接到。这代表了对网页重要性的表决形式。因此,被更频繁地链接到的网页被看作是更有价值的。当把人们带入这种考虑时,正在表决的人们是 Web 设计者,而非是内容的消费者,即用户。内容推荐的阶段 2 全是关于用户的。属于阶段 2 的三个最著名的方式是:公众分类 (folksonomy)、特征分析 / 行为分析 (profiling/behavioral targeting)、以及协同过滤 (collaborative filtering)。

[0012] 公众分类

[0013] 第一个,公众分类,代表对阶段 1 的最直接的添加。这里,允许用户自己标记内容。所以,不是网站设计者或单个的设计者,负责提出最好的关键字集合来描述该内容,公众分类让群体 (community) 来做。一旦做到这点,这些群体创建的标记本质上变成内容的一部分,并且可使用阶段 1 开发的传统的信息检索 / 搜索技术来搜索。这种方法的重要的假设是,花时间来明确标记网页的群体的子集最终产生了一种有效的、代表更大的群体看法的描述。常常并非是如此。

[0014] 特征分析 / 行为分析

[0015] 普通形式的特征分析 / 行为分析也从阶段 1 借用了大量技术。这里,基于用户在站点上的以前的行为,例如,点击的网页或购买的产品,为那个用户建立一个配置文件 (profile)。在简单的情况下,这个配置文件可基于对点击的网页或购买的产品收集。该配置文件还可利用自身的内容或元标记来尝试辨别用户的历史的感兴趣的主体。例如,如果一个用户在过去购买了许多由内容提供者标记为“恐怖”的电影,那么行为分析系统将倾向于推荐更多“恐怖”电影给该用户。这里的重要假设是,用户的历史的行为是对未来兴趣的很好的预言者。虽然有时候是真的,但是此假设往往失败和有效至少一样频繁。失败的原因在于人们展示出各种行为,取决于他们当前的兴趣、背景、和目的。例如,某人几个星期前给妻子买了几本关于吉他的书作为一次性礼物,行为分析方式可能继续推荐给他吉他书籍,即使他可能不再对这个主题感兴趣。特征分析方式常常还会考虑到用户的人口统计数据,例如年龄、性别、以及地理位置。此方式隐含的核心理念是:只要我知道关于一个用户的足够的信息,我就能准确预测他们想要什么。然而,一些基本的内省揭露了这种方式隐含的谬误。例如,我可能比其他人或机器更多地了解我的妻子。因而我是她理想的特征分析系统。然而,没有某种背景,我仍无法预测她当前正在在线寻找什么。

[0016] 协同过滤

[0017] 协同过滤是另一个以用户为中心的方式,可以说是最严格的以用户为中心。这里,基于常见的购买、点击历史或明确的评级 (ratings) 相互比较用户。例如,基于一个人在电影站点上的之前的电影评级,找到最赞同那个人的评级的其他人,并且推荐他喜欢的其他电影。标准的“购买了这个人的人也购买了那个”方式实际上是协同过滤方式的变化,其中,用户最近的动作作为识别相似的用户的基础。这种方式受到亚马逊推荐引擎的欢迎。这种方式的重要假设是,基于过去行为的用户之间的某些全局相似性度量是预测未来的兴趣的一种有用的方式。然而,这是一个有缺陷的假设。某人在工作背景中可能与他的某些的同事非常相似,例如,他们都是 Java 工程师,具有关于编程相似的兴趣,但是当在办公室外时,例如,在高尔夫球场上,此人可能与这些同事相当不同。在高尔夫的背景中,某人可能有一个非常不同的对等组 (peer group)。在全局水平上把用户分组通常更容易误导而不是更有帮助。

[0018] 阶段 2 中所有的以用户为中心的方式的另一个弱点是对喜好的明确度量或过于简单化的隐性度量的依赖。明确度量包括要用户来表明其对特定内容片段的喜好,例如,在 1-5 级上。这种方式几乎总是有偏见的,因为他们代表了非常小的人口比例。进一步地,花时间做这些评级的人们并不代表作为整体的群体。他们往往非常自以为是,或者反映了愿意花时间来发表看法的特殊个性类型。

[0019] 利用隐性意见 (implicit observations) 作为规则的这些方式查看点击或购买。用点击来评估喜好是一种有缺陷的方式,因为让某人点击结果要对引人兴趣的、可能甚至是不明确的网页上的主题和位置进行很多工作。一旦被选择来观看,它不告诉关于用户怎样寻找该内容的任何事情。另一个极端,许多系统使用购买作为喜好的度量。虽然购买是评估此的合理的方式,但是其很受限制。例如,当购得一个相机时,某人可能会在做决定前认真考虑许多产品。所有的那些信息可能对于最终购买的相机之外的其他对相机感兴趣的人是有价值的。

[0020] 发明概述

[0021] 本发明的实施方式代表内容推荐的演变的阶段 3 (见图 1)。这里,这个想法是从了解当前用户的背景开始,即:他们的意图是什么?他们在寻找什么?基于这种了解,然后找到适当的对等组,所述适当的对等组代表在此被识别的兴趣背景中与当前的用户最像的其他用户。从那里,找到那个对等组识别为与当前背景最相关的内容。

[0022] 本发明所采用的方式是以背景为中心的,或者,换一种说法,以意图为中心。用于实现此方式的技术随后介绍,并且其根本上是基于 2005 年 12 月 27 日提交的美国专利申请第 11/319,928 号中部分地描述的 UseRank 技术和亲和力引擎 (affinity engine),该申请通过引用而整体并入此处。应当注意的是,所有以前的方式,包括以内容为中心的和以用户为中心的方式,被本发明中所表现的阶段 3 包含和改进。因为本发明对在用户和内容之上的图画添加了背景的维度,所以总是可能选择忽视背景并且使用该系统来提供阶段 2 的功能性,诸如协同过滤或行为分析/特征分析。然而,基于提供全面行为指纹 (full-spectrum behavioral fingerprint) 的本发明的关键方面,即使这些以前已知的方式在其功能性方面也得以显著地改进。

[0023] 附图简述

[0024] 图 1 是逐渐引到本发明的内容推荐的演变的示意视图;



- [0025] 图 2 是示出根据本发明的一种用于基于背景的内容推荐的方法和设备的结构示意图；
- [0026] 图 3 是示出根据本发明的亲和力引擎集成的结构示意图；
- [0027] 图 4 是示出根据本发明的群体指导 Web 和群体智慧 (wisdom) 的动态的、自适应的、实时的平台的结构示意图；
- [0028] 图 5 是示出根据本发明的群体智慧平台的结构示意图；
- [0029] 图 6 是示出背景到推荐的翻译过程的示意图；
- [0030] 图 7 是示出根据本发明的虚拟公众分类 (virtual folksonomy) 的实现过程的示意图；
- [0031] 图 8 是示出根据本发明的主题吸引、主题匹配和活动性 (activeness) 的示意图；
- [0032] 图 9 是示出根据本发明的行为指纹的过程的示意图；
- [0033] 图 10 是示出根据本发明的带有结构覆盖图 (overlay) 的长尾营销 (marketing) 模型的曲线图；
- [0034] 图 11 是“群体的智慧”的拟方程 (pseudo-equation)；
- [0035] 图 12 是示出根据本发明的亲和力引擎记忆预测机器 (MemoryPrediction Machine) 的示意图；
- [0036] 图 13 是示出根据本发明的顾客门户 (portal) 的欢迎屏幕的截屏；
- [0037] 图 14 是根据本发明的顾客门户中的提升报告 (lift report) 的截屏；
- [0038] 图 15 是示出根据本发明的群体指导的电子商务的截屏, 包括根据本发明的对竞争和配套产品的推荐；
- [0039] 图 16 是示出根据本发明的群体指导的电子旅行的截屏；
- [0040] 图 17 是示出根据本发明的群体指导的营销的截屏；
- [0041] 图 18 是示出根据本发明的社会化搜索 (social search) 和群体信息弹出 / 覆盖图的截屏；
- [0042] 图 19 是示出根据本发明的群体指导的在线支持的截屏；
- [0043] 图 20 是示出根据本发明的群体内部网 / 知识门户的截屏；
- [0044] 图 21 是示出根据本发明的群体指导的媒体的截屏；
- [0045] 图 22 是示出根据本发明的对于媒体的群体指导的社会化搜索的截屏；
- [0046] 图 23 是示出根据本发明的对于媒体的群体指导的跨站点 (cross-site) 推荐的截屏和示意图；
- [0047] 图 24 是示出根据本发明的群体指导的视频推荐的截屏；
- [0048] 图 25 是示出根据本发明的群体指导的主题推荐的截屏；
- [0049] 图 26 是示出根据本发明的群体指导的相应背景的广告推荐的示意图。
- [0050] 图 27 是示出根据本发明的即时连接的示意图；
- [0051] 图 28 是示出本发明的优选实施方式的结构系统的方框示意图；
- [0052] 图 29 是示出本发明的优选实施方式中通过 JavaScript 标记和 REST 的客户集成的代码片段的方框示意图；以及
- [0053] 图 30 是示出根据本发明的 AJAX 平台的方框示意图。
- [0054] 发明详述

[0055] 本发明展现了内容推荐的演变的阶段 3(见图 1)。这里,这个想法是从先了解当前用户的背景开始,即:他们的意图是什么?他们在寻找什么?基于这种了解,然后找到适当的对等组,所述适当的对等组代表在此被识别的兴趣背景中与当前的用户最像的其他用户。从那里,找到该对等组识别为与当前背景最相关的内容。

[0056] 本发明所采用的方式是以背景为中心的,或者,换一种说法,以意图为中心。用于实现此方式的技术随后介绍,并且其根本上是基于 2005 年 12 月 27 日提交的美国专利申请第 11/319,928 号中部分地描述的 UseRank 技术和亲和力引擎。应当注意的是,所有以前的方式,包括以内容为中心的和以用户为中心的方式,被本发明中所表现的阶段 3 包含和改进。因为本发明对在用户和内容之上的图画添加了背景的维度,所以总是可能选择忽视背景而使用该系统来提供协同过滤、特征分析。然而,基于提供全面行为指纹的本发明的关键方面,即使这些以前已知的方式在其功能性方面也得以显著地改进。

[0057] 图 2 是示出根据本发明的一种用于基于背景的内容推荐的方法和设备的结构示意图。

[0058] 全面行为指纹提供针对当前技术发展现状隐性评级的重大进展,其本质上等于点击分析和购买行为。首先,他们考虑到用户的各种各样的行为,包括但不限于点击、花费在网页上的时间、滚读(scrolling)和鼠标移动、明确的动作诸如打印、电子邮件、书签使用的链接、回访频率、以及执行的搜索。第二,所有的这些行为可与其他页面上的当前用户的行为互相关,并且也可以与当前页面上的被识别的对等体的行为和群体的其余部分互相关。从此分析,该用户已经在特定内容片段中找到价值的可能性可以被辨别并且被馈入(feed)学习系统。

[0059] 本发明的另一个方面涉及对用于自动内容推荐的现有策略的无缝集成(见图 3),包括搜索引擎 301 和特征分析系统 302 以及广告服务器 303,还有用于包括推销规则及系统的人工推荐的系统 304。本发明还无缝地集成与该内容相关的其他信息资源,如产品目录 305,其可用于显示、过滤或学习的目的。

[0060] 最后,由于所学习的亲和力和收集的信息的综合性,该系统代表一个通用的智慧平台 501,在其之上可以建立许多应用(见图 4 和图 5),诸如社会化搜索 502、内容推荐 503 和产品推荐 504、Insights 505、电子邮件 506、以及专为各种应用如电子商务、电子营销等的即时连接 507。其他应用包括与投标管理系统的集成和报告,用于搜索引擎优化和搜索引擎营销 508、移动和 IPTV 509、和定制应用或混搭(mashup) 510。后文中对其中的一些应用进行描述。现在详细地描述这些关键方面中的每一个方面。

[0061] 以背景和意图为中心

[0062] 本发明所体现的内容推荐的演变中的一个核心进展是用于识别和代表用户当前的兴趣主题以及将该兴趣转变为有用的推荐和信息集的过程。此过程的步骤在图 6 中示出,并且概述如下:

[0063] 步骤 1:当用户来到一个网站,他们立即开始建立他们当前的背景。他们可能通过在该站点的搜索框内输入查询、导航到特定部分来做到这点,或甚至可能通过在外部的搜索引擎如 Google 或 Yahoo 上做通向此站点的搜索,而在到达该站点之前已经建立某种背景。所有这些信息由观察者标记(Observer Tag),即嵌入网站的一段 HTML/JavaScript,来捕获。随着用户继续通过网站移动,基于他们隐性的动作,他们还可能示出对于特定的页面或

内容片段的兴趣。基于由该观察者标记所收集的行为而确定兴趣，并且使用本发明的全面行为指纹技术 601（在下面更详细地描述）来分析兴趣。这些感兴趣的网页还进一步促成用户的背景。

[0064] 所有这些信息被储存为用户当前的背景矢量 602，其是术语和文档的混合矢量，每一个输入具有权重，反映那个术语或文档反映用户当前的背景的强烈程度。当用户输入搜索术语和 / 或点击导航链接时，增加对应输入或点击的术语和短语的矢量输入，来捕获表达的兴趣。当这些动作进一步移动成为过去，相应的输入被减少或衰减 (decay)。相似地，用户点击的或表明对其感兴趣的文档（如基于他们的隐性动作而确定的），增加了相应的矢量输入，增加的程度基于由本发明所确定的兴趣水平和确定性水平。此结果作为背景矢量，代表用户的当前背景。也有可能是，基于用户感兴趣的历史动作和历史背景，进一步增加该背景矢量。虽然在优选的实施方式中通常不会这样做，因为这种信息可能是误导性的，但是有可能基于历史行为被认为是相关的应用中的这些历史背景来较小的程度地增加该背景矢量。

[0065] 步骤 2：基于从聚合智慧 (aggregated wisdom) 604 学习到的亲和力 / 联系，扩大和改进内容矢量为意图矢量 603，聚合智慧 604 从对群体作为一个整体的长期的观察中收集。例如，用户可能已经输入关于数字 SLR 相机的查询，并且表示了对于 Nikon 网页的隐性的兴趣。在背景矢量中，对应于“数字 SLR 相机”以及“数字”、“SLR”、和“相机”的输入被较小的程度地增加，这是对应用于感兴趣的特定的 Nikon 网页的矢量输入。为创建该意图矢量，系统查看背景矢量中的术语和文档与其他术语的亲和力。例如，群体智慧可能已经发现术语“高分辨率”可能与术语“Nikon”以及与感兴趣的具体的 Nikon 网页高度相关。因而该意图矢量在对应“高分辨率”的输入处被增加。相似地，与背景矢量中的术语相关的其他文档可能基于群体智慧而被发现，并且被增加到意图矢量中。例如，Canon 相机页面可能对“SLR 相机”具有强烈的亲和力，并且某些程度上变成意图矢量的一部分。允许背景矢量扩展为意图矢量的亲和力 604 由亲和力引擎所确定，该亲和力引擎在下面描述，并且在 2005 年 12 月 27 日提交的美国专利申请第 11/319,928 号中详细描述。概括来说，亲和力引擎通过观看站点上用户群体的所有隐性行为，并通过利用行为指纹和使用等级算法来确定兴趣和联系，从而学习文档和术语、文档和文档、术语和术语、以及用户到其他用户、文档以及术语之间的关系。有效地将背景翻译成意图的能力是本发明的关键方面。

[0066] 步骤 3：识别共享对当前意图的亲和力的用户组、以及该意图的背景中展示出与当前用户最像的行为的那些用户。在 2005 年 12 月 27 日提交的美国专利申请第 11/319,928 号中，这些分别称为“专家”和“对等体”，但是这里我们将其结合成一个名字“对等体”。此对等组由用户矢量 605 代表，并且该矢量中的每一个用户输入可具有权重，表明在此背景中他对于当前用户来说是对等体的强烈程度。虽然通常未被推荐，但是系统也有能力基于预定义的属性集，诸如年龄、性别、位置或其他人口统计的变量，把对等组限制到匹配当前用户的那些用户。相似地，对等组可被限制于在一天的同一时间（例如上午、下午、晚上）访问该站点的那些用户。在某些情况下，这些属性可被用来影响用户的对等权重，诸如给予对基于这些预定义变量的用户最匹配的那些用户稍多的对等权重。本发明还通过在每一个预定义的属性组之中或之间比较行为模式，例如，最简单的例子比较被找到是有用的文档，来学习这些权重。与组间的相似性比较，更高的组内行为相似性表明该组是有个体差异的，

并且因而当影响对等组时,团体化的影响的权重更高是有理由的。可通过对用户的聚合兴趣矢量的许多相似性计算之一(例如余弦相似性)来查看使用的文档和术语相似性,来度量相似性。

[0067] 步骤 4:查看哪些文档对当前意图 606 中的所识别的对等体具有最高的亲和力。为做到这点,亲和力引擎查看多种因素组合,这些因素包括,如活动性矢量(activeness vector)所代表的一段内容的整体有用性、如术语-文档矩阵所代表的所学习到的术语和内容之间的亲和力、以及如下一步矩阵(next-step matrice)所代表的预计的导航模式,连同那些对等体,根据其对等权重,这些因素被加权,并且被聚合地用来计算具有与当前意图最高亲和力的那些文档。这些文档成为未过滤推荐 607。由下面更详细地描述,过滤现在可被用来限制或补充这些群体推荐。

[0068] 步骤 5:既然该推荐已经被识别并被适当地过滤,那么最后的可选择的步骤是向亲和力引擎要求关于每一个推荐 608 的群体信息。这些信息可与其他资产信息诸如主题、尺寸或价格结合,并被显示给用户以帮助了他们了解推荐下所隐含的群体智慧。与亲和力引擎中的文档相关联的群体智慧的许多方面可被揭示,但是在普通的实施中,我们揭示在该文档中找到有价值东西的用户的总数、将该文档与当前背景/意图相关联的对等体的数目、以及对等群体已经将其与该文档相联系的术语/短语。

[0069] 此后面的段,即,与该文档相联系的术语,被称作虚拟公众分类,因为其代表群体已经将之联系到该文档的术语,但是不像传统的公众分类,如上所述其中用户必须明确作出该联系,虚拟公众分类基于用户群体的隐性行为,由亲和力引擎自动创建。如由行为指纹所确定的,在一个或多个步骤内,被搜索或点击为导航链接并被引导到有用内容的术语,自动成为相关的。这些术语-文档连接对于提供推荐的过程,以及对于提供给用户关于与每一个推荐相联系的主题的有用的反馈,都是重要的。当这些术语被显示给用户时,可被做成可点击的,以使得用户可点击它们来提供关于他们当前意图的进一步输入。然后此信息成为他们背景向量的一部分,整个过程重复,并且可基于此新信息提供新推荐。图 7 提供了虚拟公众分类的概念图,与用于连接术语到文档的传统的方式作对比。

[0070] 时间(衰减、趋势、和时尚(fad)检测)

[0071] 在从背景到推荐的移动中,对于所有的计算,时间是很重要的因素。首先,亲和力引擎所收集的所有信息受时间衰减的影响。这意味着,例如,来自上月的信息比来自今天的信息对计算有较少的影响。这一点很重要,因为站点可能变化,例如,添加或移除了新的内容,人们可能分别地会变化,例如,他们的兴趣可能会变化,并且,作为整体的群体也可能转移兴趣,例如新的时尚。虽然对于所有信息都有默认的衰减率,但是一些信息可能更迅速地衰减,如果确定这是必要的。例如,时尚行为,诸如来得快去得快的对圣诞节产品的兴趣,可能需要被更快地衰减,以防止圣诞节已经过去很久后仍推荐圣诞节产品。为防止这点,系统定期在所有的内容上运行趋势检测系统,例如每天一次,或者对于对时间非常敏感的网站每五分钟一次。趋势检测能以多种方式做到,而一种方式是使用 Mann-Kendall 算法。其他方式包括采用最小二乘法拟合的各种回归技术。如果对于一段给定的内容,检测到了强烈的消极趋势,那么亲和力引擎中与该内容相关的信息将以更快的速率被衰减。结果是降低了亲和力引擎推荐这段内容的可能性。

[0072] 定期运行的时尚检测的另一种算法是周期性时尚检测。如果一段内容示出强烈的

积极趋势,并且在过去每隔一段时间可找到同样的趋势,那么该段内容在未来趋势的预测上可被自动增强重要性。

[0073] 在传统的网站,即使其中站点的经营者试图了解顾客行为,通常使用的技术是调查顾客并且分析顾客响应以产生报告。到该报告被分析时候,报告的结果与报告正在被评审时的实际情况不同步。例如,在传统的系统中,商务站点可能在圣诞节收集反馈并响应那个信息来在二月重新设计他们的站点。实际上,该商务站点正试图在二月销售圣诞节产品。可选择地,分析报告可被离线创建并由专家组分析,以推断趋势并确定适当的动作;花费几个星期或一个月的过程再次导致过时的站点修改。以这种方式,这里本发明的自动性允许网站实时适应所发现的趋势和时尚,提供给用户适合当前背景和时间的推荐。

[0074] 然而,除了有助于自动内容推荐的过程,本发明所使用的趋势和时尚检测算法(fad detection algorithm)还可通过顾客门户中的报告被揭示给内容系统的所有者,例如网站。在优选实施方式中,对于群体趋势和时尚的这种报告可例如被商人用来在正确的时间促销某些产品或内容,使得这些促销更为有效。以在圣诞节期间销售的产品为例,如果这种销售圣诞节后消失了,那么该商业网站的经营者将能基于与网站的群体互动来遵循销售曲线。另一方面,如果紧急的新闻报道快速驱动了对某个产品的需求,那么本发明允许商人观看需求曲线并且实时响应。因此,如果人们突然急剧地都对一段特定的内容或产品感兴趣,网站的经营者可更快地提供那种内容或产品。如果需求快速消失了,例如兴趣是基于时尚的,那么站点的经营者也能适应这一事实。

[0075] 亲和力引擎

[0076] 如上所述,并且如在2005年12月27日提交的美国专利申请第11/319,928号中更详细地讨论的,亲和力引擎学习术语-文档、术语-术语、术语-用户、文档-文档、文档-用户、以及用户-用户之间的连接/亲和力。(注意:术语如文档、内容、和资产在整个本文档中可交换地使用。在所有的情况下,这些术语是指提供信息或服务的任意类型的媒体,包括,例如但不限于网页、word文件、pdf、视频、音频文件、widget和/或产品)。在优选实施方式中,所有的亲和力被储存为稀疏矩阵和矢量。然而,还有本领域的技术人员已知的储存此信息的许多可选择的方式。虽然最终只有一个数字可被计算来代表任意两个实体例如文档和术语之间的亲和力,但是通常存在以加权和来组合的几个子亲和力(sub-affinity),来得出那个单个数字。该和的权重可能取决于背景。例如,文档具有对其他文档的至少三种亲和力:虚拟公众分类中的相似性,即,群体关联到文档的术语;用户组中的相似性,即,多少用户已经使用了两个文档;以及导航模式使用中的相似性,例如,用户示出在使用了其他的文档后在一个文档中找到价值的模式。虽然结合矢量空间模型和定制可能性的技术当前被本发明的优选实施方式所使用,但是,对每一种数据,可酌情采用多种数学技术。更多的讨论参见2005年12月27日提交的美国专利申请第11/319,928号。

[0077] 本发明的一个重要方面是,所有的子亲和力总是被双向计算。为阐明此双向性,我们描述其对来源于虚拟公众分类的术语-文档子亲和力的应用。图8是示出本发明的一些方面的结构图,这些方面涉及主题吸引和主题匹配,即构成术语和文档之间的双向虚拟公众分类连接的两个维度,以及总的活动性。在图8的示例中,提供了查询801,并且标明了在三个示例文档的一个中最终找到价值的答复的百分比。该查询涉及特定的Nikon相机,并且可以看见20%的答复在专注于高分辨率相机802的网页中找到价值,78%的答复在具体

相机的概述页 803 中找到价值,以及仅 2%的作此查询的用户在对相机的详细的规格表 804 中找到价值。作为边注,记住该价值由用户展示在给定页上的各种行为所确定(稍后详细描述)。

[0078] 有两种方式来考虑该查询和每一个示例文档之间的虚拟公众分类连接。主题吸引从查询开始,并且考虑搜索“Nikon 508X”并随后在每一个文档中找到价值的用户的比例。例如,概述页 803 具有最高的主题吸引,有 78%的用户在那里找到价值,其次是高分辨率相机页 802,有 20%的用户在那里找到价值,并且之后是规格表 804,仅有 2%的用户在那里找到价值。注意为说明起见简化了这个示例。事实上,我们也将查询分解成其组成部分,并考虑这些组成部分对每一个文档的亲合力,以及考虑亲合力引擎已经学习到的具有和这一个查询相似的意义其他术语之间的亲合力。以这种方式,给定该查询所代表的意图,主题吸引可被认为是预测用户在给定的文档中找到价值的可能性的问题。在优选的实施方式中,得到此可能性是使用概率技术的结合来完成,包括贝叶斯推理。

[0079] 主题匹配从另一方向查看,从每一个文档开始,并考虑所有的通过其他用户的行为与之相关的其他术语。所以,例如,高分辨率相机页 802 可具有连接到它的许多其他相机的名字,并且因此对输入的特定查询的关注程度比其他两种页都低,具体地,其他两种页的术语-文档连接更加高度集中围绕“Nikon 580X”。实际上,规格表 804 可变为具有术语连接最专注围绕“Nikon 580X”的文档,并且因而具有对此查询最高的主题匹配,虽然它具有最低的主题吸引。与主题吸引一样,事实上,主题匹配也将查询分解成其组成部分,并考虑不在此查询中的其他术语之间的亲合力。以这种方式,主题匹配可被视为找到具有对查询所代表的意图在主题上最佳整体匹配的文件。在优选实施方式中,矢量空间模型和修改的余弦相似性技术,与传统的全文搜索所使用的技术相似,被用来确定匹配的程度。

[0080] 考虑每一个文档对查询的亲合力的这两种方向,(即,主题吸引和主题匹配),以及最后一个因素,总的活动性(即有用性),三者被以非线性加权和结合起来,得出一个用于虚拟公众分类子亲合力的值。此子亲合力之后被与其他子亲合力(例如基于导航模式的那些亲合力)结合,并被通过对等组的透镜所过滤,得出文件对查询的最后的等级,称为 UseRank。当提供对特定的网页的推荐时,以及当考虑全面背景和意图矢量时,相似的双向技术被用来计算 UseRank。

[0081] 存在几种方法,其用于结合不同的子亲合力和其中的双向维度,以得出特定的用户和背景的文档的最终 UseRank。在优选实施方式中,在基于混合算术几何加权平均被结合在一起之前,一个或多个阈值被用于每一个维度,包括绝对和相对阈值。在再次基于混合算术几何加权平均与其他子亲合力结合之前,每一个得到的子亲合力相似性受阈值的影响。阈值的值一般是固定的,然而,加权平均的各个权重可基于由亲合力引擎返回的结果集的成功而被调节并被学习。

[0082] 全面行为指纹

[0083] 如这里所讨论,本发明的一个关键特征是处理和分析在网站的实际使用期间所做的隐性意见,例如,对比使用明确反馈的技术中的传统方式。通过科学研究证实,并从人类心理学和社会学了解,本发明的一项关键优点在于人类不擅长给予反馈,尤其如果反馈必须被明确给予时。当某人被调查时,其不具有以反馈的形式给予实际准确性的诱因。本发明的一个方面通过使用在由群体中的个人在网络上使用特定材料的期间所观察到的隐性

行为,消除了这种调查偏差。因此,本发明相信人们做什么而不观看他们说什么。本发明通过他们实际的动作、隐性的行为,观看人们的行为,并能准确诠释他们真实的意图是什么,以及他们是否喜欢或不喜欢某物。因而,对比点击动作,本发明观察行为。

[0084] 当前技术聚焦于点击,如网络链接点击。如果某人点击了链接,该点击被报告。该点击可被报告为已经导致了一个被浏览的网页,即使在此人点击的页面上没有花费太多时间。因而,不管该页面是好的或坏的,这个已知的方式无论如何不是一个好的指示。如果一个链接被放在网页上的突出位置,那么人们很可能点击它。然而,当人们到达该链接所指示的位置时,他们可能立即离开了该站点。这就是为什么浏览器上头号使用的按钮是 BACK 按钮。BACK 按钮的使用可指示对站点的喜欢或不喜欢。因此,本发明认识到,是点击后的用户动作要紧,而非点击本身。本发明跟踪点击之外的行为以确定页面是好或坏。因而,如果某人退出该网页,被认为是消极反馈,即,此人不喜欢它。以这种方式,在本发明中,点击可识别非常消极的反应。如果某人去了一个链接,追随着该链接,在那里花时间,并做了其他事,那么该行为也被跟踪。如果对等组验证该行为是一贯的,例如,相当数量的组成员对该页面反应时展示出同样的行为,那么该网页被认为是好网页。

[0085] 本发明的一种实施方式更进了一步。本发明不仅基于行为指纹确定哪些资产是有用的,其还学习与有用性相关的背景。例如,本发明可学习特定的相机页面对在高分辨率相机上示出兴趣和意图的用户非常有用,但是如果意图是袖珍相机则没有用。以这种方式,亲和力引擎有能力基于用户通过他们的隐性动作所表示的背景和意图来分辨资产的有用性。

[0086] 驱动所有的已学习到的联系(亲和力)的亲和力引擎的主要输入是用户在站点上的行为。在优选实施方式中,所有的行为被观察者标记所捕获,观察者标记即嵌入网站的一段 HTML/JavaScript,通常在文件头和文件尾模板中。虽然这是用于捕获用户行为的优选的方法,但是也可能使用浏览器插件或低级别网络流量分析仪来捕获用户行为。被捕获的行为包括:

- [0087] • 访问的网页以及以何种顺序
- [0088] • 花费在每一页面上的时间
- [0089] • 点击的链接
- [0090] • 执行的搜索
- [0091] • 在页面上滚读所花费的时间
- [0092] • 在浏览器窗口可见的页面部分,以及多长
- [0093] • 打开或关闭的页面子元素
- [0094] • 启动的媒体,观看该媒体所花费的时间,以及在该媒体上所做的明确的动作
- [0095] • BACK 按钮的使用
- [0096] • 对一段特定的内容的重复访问
- [0097] • 在该页面时的鼠标移动
- [0098] • 观看的广告和点击的广告
- [0099] • 明确的动作,例如,加入购物车、购买、发电子邮件、保存、打印、打印预览
- [0100] • 虚拟打印:在几个小时或几天的过程中经常返回的内容
- [0101] • 虚拟书签:在数小时或数天中,具有间歇的活动周期的保持打开的内容
- [0102] • 入口和出口路径

[0103] 所捕获的信息被发送回亲和力引擎用于处理。图 9 示出用于分析从各种用户接口设备所捕获的这些行为或其他隐性数据的过程。此过程的输入是用户踪迹 (trail) 901, 其包括所有的所访问的资产和在那些资产上所观察到的隐性的 (和明确的) 动作。处理这些行为有两个主要步骤, 其结合起来提供对什么内容是有用的和在什么背景中的认识。第一步是确定用户是否在一段特定的内容中找到价值 (有用性)。在概念上, 以思考模式花在页面上的时间越长, 即, 用户正在处理信息, 有用的可能性越高。思考模式可由在页面上非滚读的时间来近似, 其中一些滚读或鼠标移动已经以指定的时间范围检测到。重复访问和页面被看的百分比一般还是对喜好或有用性的好的指示者。对于每一个用户和每一段内容, 我们可创建行为矢量 902, 在该矢量中的每一个条目都代表上面列出的特征之一或预定义的特征组合。然而, 既不是所有的内容都被创建成平等的, 也不是所有的用户都被创建成平等的。我们以几种方式规范该行为矢量。第一, 我们规范用户使得行为矢量与用户群 903 其余的部分可比较。例如, 一些用户可能比其他用户阅读得慢, 影响他们花费在页面上的平均时间。一些用户可能比其他用户更多地使用鼠标。完成此规范化的一种方式, 是通过将所有的条目翻译成  $z$ -得分 ( $z$ -score), 其调整那个用户所特有的均值和标准差。

[0104] 第二个规范化是基于内容规范。基于内容的固有的或指定的属性来做到这点。例如, 花费在一个段落文档上的 30 秒很可能与花费在 10 页文档上的 30 秒具有不同的意义。因而可基于页面长度或字数来规范停留时间。相似地, 观看 30 秒长的视频所花费的 30 秒与观看 5 分钟长的视频所花费的 30 秒具有不同的意义。在许多情况中, 观察者标记有能力通过 DOM (文档对象模型) 中的信息捕获这些页面特性。然而, 当在 DOM 中不可得时, 本发明提供其他机制用于从离线目录例如产品或媒体目录收集需要的信息, 或用于允许网站设计者在页面本身中添加明确的信息, 例如, 像元标记或添加的 JavaScript 变量。例如, 在电子商务站点上, 某些页面可被定义为对比产品的信息。系统可为这些内容组的每一个独立地规范行为矢量。以这种方式, 表明喜好所需的停留时间可针对内容的类型来调节。以这种方式使得所有的行为 (特征) 是可适用的。在当前的实施中, 这些规范策略是硬编码 (hard-coded)。然而, 本发明允许插入各种机器学习技术以学习适当的规范的可能性。

[0105] 现在我们具有对于用户和内容 904 规范化的行为矢量。在当前的系统中, 预定义的规则集被用来确定此行为矢量是否代表喜欢 / 有用性。依次地, 每一个规范的特征被考虑来确定其是否满足用于指示有用性的预先指定的阈值。每一个通过的特征增加了有用性 905 的可能性。在某些情况下, 给定特征的阈值基于之前的经验而由实施系统的人预先指定并调节。在其他情况下, 这些阈值由系统动态确定。例如, 已知某些特征可呈现双峰分布, 并且阈值可被动态地确定位于两个模式之间。本发明还提供机制, 用于插入不同的机器学习技术, 以学习哪些特征是最重要的, 并且因而动态学习用于从行为矢量转换到有用性可能性的规则。

[0106] 一旦确定了有用性或有用的可能性, 分析行为的第二步是了解围绕那段内容的使用的整个背景, 包括在该使用之前所做的搜索、之前点击的链接、以及之前使用的页面。所有这种信息结合关于用户的信息来影响亲和力 906 (亲和力能以多种方式被学习和储存, 在 2005 年 12 月 27 日提交的美国专利申请第 11/319, 928 号中详细描述)。然而, 在此之前, 有一个最后的步骤: 验证 (validation)。验证是噪声滤波的一种形式, 其中, 亲和力连接被建立在例如文档和术语之间, 当且仅当足够的相似用户已经通过他们的行为证实此连



接。对于亲和力的出现,一个用户建立此连接是不够的。同样的背景(对等体)中的最少数量的用户必须已经展示出用于验证该亲和力的相似的行为和连接。我们称之为对等验证的行为。

[0107] UseRank™

[0108] 基于用户背景和意图、基于学习到的亲和力和全面的行为指纹的内容有用性的最终等级,UseRank™可媲美因 Google 而出名的非常流行的 PageRank 策略。在 PageRank 中,基于链接到它的其他页面的数量,给予每一个网页一个值。另外,来自本身具有高值的网页的链接将 PageRank 甚至提高得更多。这种民主(democratic)的策略在让用户基于他们的 Google 查询到达有用的网站中是非常有效的,但是一旦用户开始在网站本身中寻找进一步的信息时一般就行不通了。对此的主要原因是在一个网站中的多个页面,尤其是那些带有长尾内容的网页,不管在外部或内部都没有被紧密地链接到。在内部,链接由网站的结构唯一确定,并且更多的链接不一定意味着更多的价值。另外,在此类站点上的许多页面不是 HTML 格式的,例如 PDF、Word 文档、或视频,并且根本不链接到其他内容。

[0109] 基于网站上的用户行为,UseRank 方法学减少了所有这些问题。代替依靠 Web 设计者对文档的链接,其依靠 Web 用户对文档的用法,Web 用户在任何网站上都是丰富的信息来源。用法是学习什么内容是最有价值的真正民主的方式。事实上,PageRank 的设计者认识到他们的方式是对用户活动的近似。考虑到与跟踪用户在整个 Web 上的行为有关的隐私问题,Google 选择基于 PageRank 来近似用户价值是有意义的。在 UseRank 的优选实施方式中,我们仅跟踪配备有观察者标记的站点上的用户行为,并且总是以匿名的方式,除非由部署观察者标记的网站以其它方式配置。长尾

[0110] 本发明的另一个重要方面出现于亲和力引擎所隐含的背景驱动方式。以内容为中心或以用户为中心的许多之前的系统都有一个问题,其中仅仅最受欢迎的产品/内容被推荐。因为这些系统缺乏对用户当前背景/意图的深刻了解,他们尽了最大的努力。本发明提供以背景为中心的方式,其允许亲和力引擎将其关注点窄化到匹配当前用户的背景/意图的用户(对等体)和内容的子集,即使该意图在大的事物体系中不是非常受欢迎的。这就是已知的产品的长尾,即,这些产品总的来说不是特别流行的,但是极为重要,因为他们强烈满足群体中小而重要的子集的需要(见 C. Anderson, *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion 出版社(2006))。本发明有能力识别随同有关的对等组和相关的内容的长尾兴趣,并且因而可在适当的时间推荐那些重要的长尾产品/内容(见图 10)。这点对于产品/内容提供者来说至关重要,因为经常是这些长尾产品对于群体最有用,并经常会导致公司本身获得最高利润或利益。

[0111] 因此,本发明使用志趣相投(like-minded)的对等体作为使能者来沿着长尾定位小目标片段,而不是使用在现有技术中已知的个人问题个性化或历史兴趣。本发明提供背景化定位(contextual targeting)。给定背景中的人们的需要在志趣相投的人们中通常类似。本发明的洞察是,人们具有数以百计的特征,如果不具有数以千计的特征的话。这些特征没有交叉连接到现实,所以没有背景时,很难预测个人的使用水平。一旦个人是在特定的背景中,例如,在志趣相投的对等体中,该人表现得与组背景中的其他每个人类似。因而,如果某人正在访问高尔夫器械站点去买高尔夫驱动器或球棒,那么他表现得与其他打高尔夫球的人相似。此人的政治偏见或他们的文化背景是什么是无紧要的。个人的这些方面没

有相关性,因为此人基于组成员身份的目前的背景是更有关的。

#### [0112] 群体的智慧

[0113] 本发明代表一种利用隐性的群体智慧来创建自适应的网站和其他信息门户的新方式。在其著作 *The Wisdom of Crowds* 中,James Surowiecki 解释了一大组平均个人的集体才智几乎总是超过了专家的才智。为说明此概念,他使用来自县集市的一个示例,其中一组集市参与者尝试猜测一头奶牛的重量。一组所谓的专家例如屠夫、奶农等也做猜测。最后,专家的猜测都偏离了很大的数目。然而,非专家的群体所猜测的平均重量与奶牛的实际重量相差在一磅之内。Surowiecki 继续示出如何在从股票市场预测到民主统治的每一件事中看出相似的现象。个体行为者的组可集体地展示出集体才智水平,甚至超越个体行为者自身的总和,这个观念在生物学和人工智能领域被称为突现行为 (emergent behavior) 和集体的才智。大自然当中存在无数的示例,其中非常简单的个体动物,例如蚂蚁或蜜蜂,在满足组的需要方面,集体展示出非凡的才能和机智,例如觅食或筑巢。

[0114] 在网站设计的背景中,Surowiecki 的专家是 Web 设计者。这些专家试图对哪些网页应该链接到另外哪些网页作出正确决策,以使得访问者最能找到他们所寻找的。Web 设计者还可能在调节搜索结果以匹配其访问者预期的需要的努力上花费许多时间。在此背景中的群体是来到此站点的大组网站访问者。一般地,群体对网站的组织没有直接的影响。他们保持沉默。本发明给此群体出了个主意,使得他们的集体的动作,即,所表达的看法的形式,被收集,并且通过他们对所推荐的内容的直接影响来自动驱动对网站组织的决策。因此,本发明出于创建有用网站的目的,深入了解群体的智慧。虽然在优选的实施方式中,群体智慧的影响被划分为网站中的推荐和社会化搜索的具体区域,但是容易想象扩展的执行,其中网站的整个组织、其所有的链接和菜单,例如,最终由网站访问者的行为驱动并适应于网站访问者的行为。以这种方式,由多个相互连接的站点所组成的整个互联网可开始演变并适应为一种最能满足互联网用户需要的形式。

[0115] 图 11 提供群体智慧的拟方程,示出背景 / 意图如何以广义的方式被融合来扩展群体智慧的概念,如本发明的一种实施方式所示例的。在此拟方程中,每一个用户的隐性表决根据当前用户背景和用户在其中作出表决的背景之间的相似性被加权。此方程可被进一步扩展来融合用户之间的相似性。当所有的背景都是相同的,那么此方程正确简化为一个简单的平均。记忆预测机器

[0116] 亲和力引擎的动作还可被认为是一种记忆预测机器的形式。近来的从认知科学的学习教授了:大脑被构造成对在记忆的目标区域内的对象与概念之间或之中的联系进行捕获和编码,并以从抽象到具体的层次来组织那些联系。基于之前的刺激造成的记忆,对新的刺激做出响应并预测其结果,并因而编码所学习到的联系。在网站的环境中,相似地,亲和力引擎通过观察用户、对象和背景之间的互动,来学习他们之间的联系,并且在其存储器中记住那些模式,并有层次地储存。当用户展示之前已经由亲和力引擎学习到的背景时,亲和力引擎指明了存储器中的适当的联系,这触发了对最满足当前用户和背景的需要的那些对象的预测。图 12 是其功能作为记忆预测机器的亲和力引擎的结构示意图。

#### [0117] 与外部系统的无缝集成

[0118] 本发明的一个重要的方面是其与其他推荐系统、政策和信息来源无缝集成的能力。一些示例在图 3 中示出。

### [0119] 搜索引擎

[0120] 当用户在以本发明实施的站点上执行搜索时,搜索术语被发送到亲和力引擎进行处理。亲和力引擎基于从过去的群体行为学习到的亲和力,确定给定的那个搜索的背景下最有用的内容集。然而,亲和力还可考虑来自外部来源的看法,例如搜索引擎。搜索引擎 301 有它自己的亲和力引擎可接受的推荐集。来自搜索引擎的相关性评级可与亲和力引擎中所体现的信息结合,来产生单个的统一的推荐集(在 2005 年 12 月 27 日提交的美国专利申请第 11/319,928 号中更充分地讨论)。XML feed 是用于本系统与搜索引擎接口的一种典型的方式。

### [0121] 产品目录

[0122] 本发明还提供用于自动输入产品或媒体目录 305 的机制。这样做的优选方式是使用系统可访问的 XML feed。目录可用三种方式来使用:

[0123] 1) 来自目录的信息可随着推荐一起被显示,如摘要或价格;

[0124] 2) 来自目录的属性可被用来过滤结果集,例如,网站设计者可能想要把推荐限制为仅为网站的特定部分的产品或 PDF。

[0125] 3) 在上述行为指纹过程中,出于规范的目的,内容分类可被用来对内容分组。

### [0126] 广告服务器

[0127] 本发明的另一方面在于它是内容无关的。系统可观察进行的行为并推荐网页、图片、视频、文档、博客、下载以及甚至广告。在推荐广告的情况下,经常有必要与广告服务器集成。广告服务器 303 可提供与如上所述的搜索引擎和目录系统都相似的功能性。本发明以与那些接口相似的方式与广告服务器接口,通常通过 XML。

### [0128] 推销系统

[0129] 推销系统 304 在一般情况下为网站所有者提供一种机制以影响什么产品/内容被推荐给用户。在本发明中,优选地,群体是主要的推荐驱动器。然而,有些时候,所有者有必须被满足的具体需要,但与群体所表示的兴趣无关。例如,在特定的背景中,所有者可总是选择首先推荐一特定的产品或一段内容,例如,可能有产品的促销。所有者也可能希望影响群体来购买具有更高利润的产品,或防止某段内容被推荐。本发明有两种方式来实践这种影响。首先是通过作为本发明的一个组件的定制规则接口。这里,网站所有者可登录,并从多种预定义的规则类型中选择,规则类型包括锁定,即,迫使一个特定的推荐显示;列入黑名单,即,防止一个给定的推荐或一类推荐显示;和激励,即,增加一个推荐或一类推荐显示的机会,但是仍然实践群体智慧。对于给定的搜索术语,对于给定的一类用户,在特定的时间范围,所有的这些规则类型可被全局应用或仅在一个给定的背景中应用,例如在一个特定的页面上。另一种规则类型允许所有者充分实践群体智慧,但是影响推荐的目标。例如,亲和力引擎可被告知来推荐用户最可能购买的产品,而非关注于推荐有用的产品。另一个示例,亲和力引擎可被要求来推荐最可能引导用户到特定的目标页面集的内容。除了定制规则接口,本发明提供用于以 XML 形式从外部推销应用中输入规则的机制。当规则类型存在于外部系统中而在本发明中还不存在时,则定制规则插件可被设计用于所发明的系统,其匹配外部规则类型所期望的行为。

### [0130] 用户特征分析

[0131] 亲和力引擎可与之集成的另一类型系统是用户特征分析系统 302。在其最简单的

形式中,特征分析系统是与用户相关的属性集,例如人口统计属性。这些属性可在推荐请求时被传给亲和力引擎。如前所述,亲和力引擎可在识别对等组时考虑这些属性,其然后影响返回的推荐。更复杂的特征分析系统要求与其他的外部组件相似的 XML 集成。这里,亲和力引擎可能需要在对等识别过程期间动态地联系外部特征分析系统。

#### [0132] 优选实施方式

[0133] 在本发明的优选实施方式中,推荐系统在软件即服务 (Software as aService, SaaS) 模型被实施,以帮助网站访问者找到他们喜欢或需要的产品或内容的自动建议。图 2 是用于基于背景的内容推荐的方法和设备的结构示意图。图 2 中,用户与网站的互动产生隐性的突现行为。这种用户互动包括页面转送、链接、条目踪迹、查询、页面大小、鼠标移动、对等体、消极经验、虚拟书签、花费的时间、虚拟打印、退出踪迹,诸如此类。根据本发明在推荐/亲和力引擎中处理这些行为,导致以社会化搜索和导航指导的形式的自动的内容推荐和产品推荐,还提供实时反馈给商人,例如关于访问者俱乐部 (Visitor Club),以及识别内容缺口 (gap) (其中每一项下面都更详细地讨论)。下面的关键过程包括推荐系统:

[0134] • 捕获隐性行为——隐性网站行为 201 作为推荐系统的输入在客户端由嵌入网站页面的小片段 JavaScript 代码捕获,例如,通过网站模板。然后该行为被发送到远程推荐引擎,在那里使用这里描述的全面行为指纹技术来将其处理。

[0135] • 提炼集体的智慧——推荐/亲和力引擎 202 处理所有进来的信息以识别网站访问者中出现的意图,并捕获群体的集体智慧。推荐引擎识别并学习用户、内容和术语之间和之中的亲和力,识别一贯的行为模式和消除噪声。亲和力引擎还实时认出并适应不断变化的时尚和趋势、以及周期性的行为模式例如季节性模式、和其他的用户兴趣的转移。

[0136] • 传递推荐结果和搜索结果——当访问者到达页面时,推荐引擎被传送了关于用户当前背景的信息,要么由称为推荐标记的客户的 JavaScript 片段传送,要么由网站服务器传送,并且请求适当的推荐。推荐引擎使用此申请中所描述的技术将用户背景翻译成意图,识别适当的对等组,并最终返回一组内容推荐 203,这些内容推荐基于其被计算出的 UseRank™ 而排序。这些推荐之后被显示在用户浏览器的网页中。这些推荐可采取导航链接或增强的搜索结果的形式。当传递推荐结果时,推荐系统可另外联系其他外部服务,诸如网站的全文搜索引擎、产品目录或广告服务器 (见图 3)。

[0137] • 顾客门户——在优选实施方式中,网站的所有者也可通过称为 Insights 204 的顾客门户访问由推荐引擎所学习到的信息和模式。顾客门户提供能力来配置推荐系统和添加推销规则、以及观看由提供关于群体对网站使用的信息的一组报告以及系统所学习到的行为模式和亲和力。图 13 示出在优选的实施方式中的顾客门户主页的截屏。通过使用此门户中的链接可访问许多不同的报告和配置。例如,本发明允许产品/内容缺口检测,其中缺口基于站点上的用法模式被检测,并且帮助站点所有者在缺口被检测的那些区域中引进新的产品或内容。另一个示例,顾客门户还提供一组报告给网站所有者,以跟踪推荐引擎正带给网站的持续的价值。图 14 示出跟踪在网站上由于推荐的存在而收入提升的 A/B 报告的截屏。发明人始终如一地看到收入提高了 20-30% 或更多。在门户中可得的其他提升报告包括改善网站中的页面浏览、投入时间、站点粘着性 (stickiness)、平均订单价值、和到指定“商业目标”的用户转变。

#### [0138] 应用

[0139] 下面的讨论涉及上述技术的不同应用：

[0140] 图 15 是示出群体指导的电子商务的截屏。在图 15 中，呈现了产品推荐的 5 种不同的能力：

[0141] 1) 基于志趣相投的对等体关于他们的产品考虑的亲和力，即，如前所述的观察到的全面的行为指纹，出于比较购物和向上销售 (up-selling) 的目的，示出比较的产品，也称为相似的或竞争的产品。基于考虑的对等推荐推动了更高端 / 特殊的产品，并为站点产生更多的收入和利润。

[0142] 2) 附属产品也被示出，用于相关的交叉销售 / 非竞争的产品例如配件，以增加总订单量。

[0143] 3) 本发明还示出最流行的产品，如整个站点上或某个分类中最具吸引力的产品。

[0144] 4) 本发明还产生意图驱动 (intent-driven) 的登陆页面 (landing page)，也称为 AdGuide 和 Site Concierge，具有基于访问者在标准的 Web 搜索引擎如 Google、Yahoo 或 MSN 上搜索什么的推荐。如果访问者在 Google 上搜索“Viking Cooktop”并登陆了顾客站点，AdGuide 会将最好的 VikingCooktop 推荐动态地提供给登陆页面上的访问者，而非示给他一些无关的事物。

[0145] 5) 本发明还提供基于上述的对产品的群体过滤的搜索。

[0146] 对于对等驱动的推荐，本发明自动识别具有独特的兴趣的小型的和大型的人口部分，并引导个人到流行的、竞争的和附属的产品。本发明提供用户意图和产品映射，其将购物者的意图翻译成对等验证的产品和品牌。发明者已经基于独立的研究找到 7 倍的转换能力。本发明提供时尚和趋势检测，并实时识别和适应购物者兴趣的季节性的、促销的和其他转移。本发明还允许产品缺口检测，其中产品缺口被检测并且帮助网站所有者在缺口被检测到的那些区域引进新的产品或内容。本发明用实时顾客反馈协助推销。其与现有的推销、促销、分段和搜索一起工作，并扩大了产品或产品家族。本发明实时地提供对由无形的群体所产生的净收入的内置的同时的 A/B 测试和报告。

[0147] 图 16 是示出具有旅行示例的群体指导的产品搜索的截屏。在图 16 中，用户已经查询纽约市的旅馆。由于此网站的相似的用户出于相似的目的，并通过观察用户怎样使用返回的信息，对此人推荐关于在纽约市中选择旅馆。基于志趣相投的对等体的兴趣和意图而非通过简单匹配关键字和元数据（这个情况下是“纽约旅馆”）来示出许多不相关的结果，此搜索结果更加有效和相关。

[0148] 图 17 是示出群体指导的营销和在线消费勘察 (online lead generation) 的截屏。如同电子营销应用一样，五种不同的能力被提供给营销网站。

[0149] 1) 本发明此特征的方面包括涉及 UseRank™ 和隐性学习的社会化搜索。这种方式是意图驱动的、自适应的，并利用了隐性公众分类、群体术语，例如他们使用的链接文本以及他们输入的搜索查询，如上所述。本发明支持音频、视频、二进制代码和所有其他内容类型，因为它不必自身解析内容，而典型的搜索引擎需要这么做，因而仅限于文本内容和元数据。

[0150] 2) 本发明还基于志趣相投的用户的隐性的行为反馈来提供相关的或相似的内容。本发明基于背景和意图，通过有用性和切合目标性将推荐分等级。

[0151] 3) 本发明还提供最流行的分类，其提供在整个站点或站点的子部分中具有最大价

值的内容。此促销信息导致商业目标转变的增加,如试用/下载/注册转变,并可在网站或主题的水平下被提供。

[0152] 4) 提供了下一步特征,其涉及导致顾客转变的常见的接下去的步骤。下一步可能是商业过程中的任何一个下一步或与该过程有关的将要阅读的正常的下一个内容集。

[0153] 5) 相似的意图驱动的登陆页面,如同上述的电子营销应用一样,对于营销网站是可用的。

[0154] 图 18 是示出根据本发明的社会化搜索的截屏。在该示例中,对储存设备进行搜索。用户被提供有信息,其是关于信息是否切合目标和多少访问者、以及流行评级和群体等级,即 UseRank™。切合目标的值也与结果一同被提供。进一步地,基于亲和力引擎中的信息,结果由有用性分类。作为之前和之后的比较,还提供了来自 IR 或全文搜索引擎的初始的关键词等级。在此示例中,如果没有使用 UseRank™,主要的储存产品的这段好信息可能作为等级 87 而被掩藏了,即第 9 搜索结果页面上的第 7 个结果,因而没有人能找到和使用它。UseRank™ 已经将它移动到搜索结果的第一页面上的第三位。在接口水平,如果那里不使用本发明,则原始结果链接被提供来给予用户初始结果。这是另外一种看见 UseRank™ 的之前和之后的影响的方式。

[0155] 图 19 是示出群体指导的在线支持以促进顾客群中的自助服务和满意的截屏。此页面示出顶级问题解决方案,包括最有帮助的文章、最常见的问题解答、和帮助论坛、以及博客文章。此页面还示出相关帮助,包括关于这个主题的有深度的文章、可选择的解决办法、以及,最后地提供了提示。进一步地,提供了常见的随后步骤,包括解决方案、常见的下载、以及附加的支持信息和联系信息的自然的随后步骤。最后,基于 UseRank™ 和隐性学习,本发明示出社会化搜索,其是本发明的此实施方式的关键方面。因此,社会化搜索是意图驱动的和自适应的,利用了隐性公众分类,并支持所有的内容类型。对于这种使用的其他好的支持应用包括知识管理 (KM) 门户、社区论坛和讨论线索 (discussion thread)、聊天、维基和开发者的网络等等。

[0156] 图 20 示出群体指导的内部网和知识门户。在此示例中,社会化搜索被用于内部网中。该搜索返回相关页面,其包括关于这个主题的相似的页面,这些页面由过去的对等体经验所证明,并且启用偶然发现。进一步地,该搜索过滤掉低价值的内容,本发明还提供最流行的搜索结果,其包括具有最大价值的内容、公司内的顶级应用、和部门所特有的信息或跨公司的信息。最后,本发明提供随后的步骤,包括人们从此页面去满足他们的目的、在商业过程中的下一步、以及关于该主题的将阅读的下一个内容集。外部网,例如医患门户、经销商网、顾客门户,是其他应用示例。

[0157] 图 21 示出在媒体站点上所实施的推荐,并且图 22 示出在媒体站点上所实施的社会化搜索。图 23 是示出群体指导的媒体和跨站点的背景的推荐的截屏。相似的文章和博客基于相似的背景被推荐。如果你正在阅读泰格伍兹高尔夫故事,基于其他相似的对等兴趣,其他三个相关的泰格新闻被供给相同的访问者。访问者得到更多的内容并满意。此站点得到更高的 CPM 和 CPA,这两个都是从广告的角度对访问者转换率和站点价值的当前工业测量。也被 Baynote 称为 BrandRank™ 的 UseRank™ 还可被用来将关于媒体的广告和产品重新分等级,而非随机地或基于关键字匹配放置广告和产品。对等体所验证的广告具有明显更高的 CPM 和 CPA 值,并且因此增加了媒体站点的收入。另外,这些不同的来源所包含的

信息被用于亲和力引擎,其发现不同站点中的跨站点亲和力。结果是提供切题的内容和产品。

[0158] 图 24 是本发明被用来推荐视频内容的示例。如此申请所述,本发明是内容无关的并有能力完全基于隐性用户动作来学习术语、用户和内容之间的亲和力。

[0159] 图 25 示出隐性主题云 (Implicit Topic Cloud) 的截屏。本发明的另一个应用是收集和显示群体词汇或术语。表面上,主题云看起来与明确的加了标记的云类似,如 Deli-cio-us、社会书签等。但是关键的不同在于隐性。隐性主题云是基于站点上实际的用户行为建立的,即,他们所输入的查询和他们所使用的链接文本。行动胜于雄辩。隐性主题云反映在站点访问者的 100% 的兴趣中,不是愿意给事物加标记的少数的大声的访问者。沉默的大多数是收集真实群体智慧的关键。

[0160] 图 26 示出本发明和广告服务器之间的集成的结构视图。为将最好的广告提供给用户,集成系统首先联系现有的广告服务器,以得到向给定的人口统计和其他信息的用户显示的一个可接受的广告列表。然后此列表被发送到亲和力引擎,其基于广告、用户和内容之间的所学习到的亲和力,选择最好的广告显示给当前用户。这允许所提供的广告对于用户以及用户当前的背景和意图都是适当的。

[0161] 除了社会化搜索和在线内容 / 产品推荐,还有许多方式来利用由推荐引擎所提炼的群体智慧。许多应用可被建立或修改来通过对推荐引擎的 API 而直接深入了解此智慧。图 4 是示出本发明的作用是群体智慧平台的结构图。群体智慧平台还可包括这些条目,如背景化电子邮件营销 (contextual email marketing)、移动应用、IPTV 系统、SEO/SEM 应用和定制应用中的混搭。在这些系统中,来自推荐引擎的输入可被用来动态确定最适当的信息和 / 或信息组织以呈现给用户。这种平台是建立在包括如 AJAX 和 REST (下面讨论) 这些特征的 Web 服务上的。因此,本发明是建立在新版 Web 2.0 基础之上的,其包括对无形群体的真实了解、从志趣相投的对等体聚集信息、作出内容推荐或产品推荐、提供现场的社会化搜索、隐性的基于群体的报告,诸如此类。下面的描述给出关于他们怎样被部署的细节:

[0162] 电子邮件推荐:营销使用大量的电子邮件与顾客和可能的顾客联系。但是电子邮件活动的内容一般由营销人员确定。本发明通过众包 (crowd-sourcing) 给志趣相投的对等体而提高了电子邮件活动的质量。电子邮件的内容由流行的或是与志趣相投的电子邮件读者感兴趣的事物相似的事物所确定。营销商简便地促进了让一个顾客隐性地将内容推广给另一个顾客的过程。

[0163] 移动推荐:由于移动设备如手机和个人设备的小尺寸,移动推荐的质量要求甚至比 Web 更高。本发明的志趣相投的对等体和意图驱动的方式好于传统的基于协同过滤、页面浏览量和人口统计或购买数据的方式。意图 / 背景驱动的对等体预测人们的需求远远好于竞争方式。

[0164] IPTV 推荐:对于使用数字 TV/HDTV 来观看互联网所传送的电影和视频节目,相似的推荐是有效的。有数以千计的电影和 TV 节目选择,像“你还可能喜欢这些电影”的特征对于向上销售和交叉销售电影和 TV 节目是极为重要的。背景驱动的对等推荐是对于喜欢该推荐的观众最有效的方式。

[0165] 即时连接:图 27 示出即时连接系统的结构视图。由于本发明可识别志趣相投的对等体,除了隐性地收获他们的集体智慧和对个人做出对等推荐,本发明还可将个人与志趣

相投的对等组连接起来并让他们交流信息、知识和经验。这同彼此不认识的志趣相投的对等体的现实世界的经验相似。想象你在 Best Buy 购买 TV, 你看见其他的志趣相投的对等体也在购买 TV, 你可以向他们询问他们关于某些产品的经验。

[0166] SEO/SEM 推荐: 本发明还用于在登陆站点的时刻指导网站访问者。此特征被称为“AdGuide”, “Site Concierge”, 和“Site Maitre’ d”。本申请检查人们已经在 Google、Yahoo、MSN、ASK 和其他主要 Web 搜索引擎上搜索的查询, 然后使用这些查询作为用户意图的代理, 来为访问者在其登陆该站点时, 在无论是预先设计的登陆页面 (SEM) 还是正常的搜索结果页面 (SEO) 上, 显示正确的产品或内容。例如, 如果访问者在 Google 上搜索“Viking Cooktop”并登陆 Baynote 顾客站点, AdGuide 基于志趣相投的对等体的集体的兴趣, 在登陆页面上将最好的 Viking cooktop 推荐动态地提供给访问者, 而非通过简单地匹配关键字给他显示一些不相关的或不太重要的内容。

[0167] 基于亲和力引擎所学习到的亲和力, 本发明还有能力提供关键字推荐系统, 其提供建议给营销商, 在该系统上关键字应该在外部的搜索引擎上被购买, 作为他们的 SEO/SEM 努力的一部分。例如, 本发明可建议群体最常使用来描述网站上的内容的那些术语, 以及最有可能导致站点中有用的内容的那些术语。此外, 假设特定集合的关键字已经被购买了, 本发明可建议群体所使用的具有对现有集合高的亲和力的其他词语。要么是通过顾客门户中的报告, 这些推荐可被提供给 SEO/SEM 商人, 或者这些推荐可直接与外部 SEO/SEM 系统集成, 以动态地和自动地竞标推荐关键字。

[0168] 博客、论坛、和讨论线索推荐: 本发明可用于群体产生的内容, 诸如博客、论坛、讨论、播客 (pod cast)、视频内容等等。由于涉及的内容数量, 以正确的顺序呈现正确的内容比专家产生的网站甚至更重要。使用志趣相投的对等体用于推荐和社会化搜索是极为重要和有效的。这些网站可以是面向公众的博客站点、或要求登录的站点如伙伴门户或医患门户、或开发者的网络、或内部网。诸如“相似的内容”、“附属产品”、“接下来的步骤”、“最流行的”的特征和社会化搜索是非常有用和有效的。

[0169] Insights (访问者俱乐部、内容缺口等): Insights, 顾客门户, 早些时候已经在本申请中部分地讨论了。虽然本发明的主要使用是作为自动推荐系统, 但是有时候站点管理者可能想要基于亲和力引擎所学习到的知识, 来修改推荐系统的行为或观看报告。Insights 目前通过它的 UI 提供三个主要的功能分类: 配置、管理和报告。配置使得站点管理者充分配置完整的功能和部署所需的推荐引擎的所有方面, 包括与信息的外部来源如全文搜索引擎和广告服务器集成。管理使得站点管理者能够以整体的方式或对于规定的背景所特有的方式 (例如一个特定的搜索术语或页面位置或用户), 来修改社会化搜索和推荐的行为。例如, 管理者可创建产品 / 内容促销, 以在设置的时间段凌驾于群体智慧之上, 可将产品 / 内容加入黑名单而不被推荐, 人工修改或提升规定的产品 / 内容类别, 或建立商业规则例如限制适用于特定的背景中的推荐的产品 / 推荐的类别。报告提供关于由亲和力引擎所提炼的群体智慧的各种视图。虽然报告包括在传统的分析中可找到的基于点击的和基于购买的信息的子集, 但是它们超越了这些, 提供了许多关于群体智慧的细节, 该群体智慧通过由亲和力引擎所学习到的全面的行为指纹和连接被提炼。示例报告讨论如下。

[0170] 内容缺口报告使用亲和力引擎来分析未被站点上现有的内容满足的那些用户兴趣。使用全面行为指纹技术, 并通过分析本申请早些时候讨论的群体行为, 亲和力引擎可识



别群体兴趣。然后亲和力引擎可分析所有的情况,其中特定兴趣由用户和该用户随后的行为表达,包括内容的使用和提取的价值水平、以及随后的搜索或退出。如果大部分的具有特定兴趣的用户群未能找到有用内容,那么该兴趣被指定为内容缺口。由亲和力引擎所推测的缺口达到的程度也可被报告。然后管理者可采取必要的动作以通过添加新内容来补救该缺口,并且之后继续使用内容缺口报告来查看缺口存在的影响。

[0171] 访问者俱乐部报告提供关于由亲和力引擎所学习到的亲和力的最全面的观点。站点的访问者基于共享的兴趣被隐性地分组。每一个兴趣组或访问者俱乐部,随同该俱乐部中的访问者的数量、他们总体的活动水平、描述该俱乐部的术语的虚拟公众分类、以及对于在该俱乐部的兴趣背景中对俱乐部最有用的内容/产品,一起被显示在界面中。另外,所有的俱乐部特性可随着时间的推移被绘制,以了解兴趣随着时间的推移的趋势如何。对于成员身份的重叠,俱乐部还可相互比较。此报告把在用户、内容/产品、和术语/背景之间探索所学习到的联系的能力提供给管理者。之后这些联系可在 Insights 的管理部分中驱动促销或其他推销活动。访问者俱乐部信息还可用于驱动所发明的系统的背景之外的不同的商业决策。

[0172] 结构

[0173] 推荐服务器

[0174] 图 3 是示出根据本发明的亲和力引擎集成的结构示意图。图 3 中,在特定的内容站点中展示可观测的行为(如当前位置和/或搜索)的当前用户,基于之前的动作或用户属性,来访问亲和力引擎,其包括资产/产品、用户和术语,全部被基于亲和力安排归类。提供关于产品、分类、信息、媒体等诸如此类的资产推荐给用户。亲和力引擎由站点所有者填充,站点所有者提供管理性输入和施加管理性影响。信息由 XML/直接输入提供。管理性输入和管理性影响过滤、限制和催化亲和力引擎,施加推销规则,以及表达目的,例如,可能使用或可能购买特定的产品。资产信息还可以 XML 信息的形式(例如产品目录或全文搜索工具)被提供。亲和力引擎从访问者群体收集隐性意见,以对当前用户作出这种资产推荐。这些隐性意见包括资产动作,例如页面上所花费的时间、滚读、鼠标移动或点击、页面互动、和虚拟打印或书签;资产导航和资产导航模式,包括例如进入路径、退出路径、所使用的连接、重复的访问、以及频率访问、后退按钮的使用、以及推荐使用;和访问者群体搜索,包括输入和/或重新输入的术语、点击的结果、使用的结果、和随后的导航。

[0175] 图 28 是示出本发明的优选实施方式的系统结构的方框示意图。此结构的各个方面的更加详细的讨论提供如下。为了更详细地了解图 28,必须参考申请人的共同未决的 2005 年 12 月 27 日提交的美国专利申请第 11/319,928 号,其通过引用而被整体并入此处,并且其更详细地描述了用于有关本发明的使用的优选的系统。

[0176] 此结构由服务器场 20、顾客企业 22、和用户浏览器 21 组成。用户浏览器配备有扩展 23,并访问在顾客企业 22 的顾客服务器 25 以及经由负载均衡器 27 访问服务器场 20。目前使用 HTTPS 协议实现与服务器场的通信。用户访问顾客服务器合乎于企业协议。浏览器扩展 23 更详细地讨论如下。

[0177] 显要的与本发明有关的是提供故障保护。扩展 23 以及企业扩展 24 被构造成使得,如果服务器场 20 未以成功的方式响应,那么该扩展被关闭,并且企业和浏览器以正常的方式互动。本发明的特征仅在服务器是活动的并正确执行它的操作时提供。因此,服务器的

失败无论如何不会损害对于企业用户的企业操作。

[0178] 如上所述,扩展 24 还被提供用于企业,其经由 HTTPS 协议,在服务器场 20 与负载均衡器 27 通信。

[0179] 企业还包括助手(helper)26,其经由代理 31 使用 HTTPS 协议与服务器场通信。代理从企业检索日志信息并将其提供给日志分析器 28,其产生的结果被呈现给用法库 29。为了下载信息到亲和力引擎 32 中,代理还与企业可访问的任何数据来源通信。这些信息包括目录数据库信息、使用配置文件数据、身份验证数据、以及对于确定或过滤推荐和搜索结果的目的有用的任何其他形式的数据。

[0180] 信息经由不同的分配器 30 在亲和力引擎 32 和浏览器和企业之间交换。自此,浏览器本身提供意见给服务器,并响应于来于此的搜索或推荐请求接收显示。可选择地,顾客服务器 25 可使用 REST,通过 Baynote 扩展 24,直接从亲和力引擎 32 直接请求搜索结果或推荐结果。

[0181] 本发明的一个关键特征是亲和力引擎 32,其包括多个处理器 33/34 和配置管理工具 35。在本发明的操作期间,一种形式的信息,也称为智慧,被收集在智慧数据库 36 中。

[0182] 客户端集成

[0183] Baynote 扩展 23 的优选实施方式使用定制建立的基于标记的 AJAX 结构,用于从客户收集网站行为,以及将推荐提供给客户。图 29 提供与客户集成的 JavaScript 代码片段。图 30 是示出根据本发明的 AJAX 标记平台的方框示意图。图 30 中,网页被显示在客户机上,并包括特定的脚本。在此示例中,来自系统的所有的文件和结果被作为 <script> 元素动态注入 DOMS。可信的 Web 服务器上的协调器提供诸如故障保护机制、共同的代码、政策、处理机(handler)、和通信工具这些事物。系统服务器提供诸如心跳(heartbeat)、共同的 JavaScript、政策、处理机如观察者和向导、以及亲和力引擎这类单元。图 30 中,协调器是位于可信网站上的一段代码。在目前优选实施方式中,协调器驻留在顾客的网站。因此,有一个文件作为协调器的角色来服务,并驻留在特定的顾客网站上,然而它可以驻留在任何其他可信的网站上。本发明的此方面由于故障保护的原因,涉及可信 Web 服务器的观念。重要的是确定如果所发明的系统崩溃了,它不会使客户网站崩溃或阻碍客户网站。因此,协调器代码被提供作为故障保护程序。它检查故障保护程序,来确保服务器的网站崩溃后的正确的恢复。

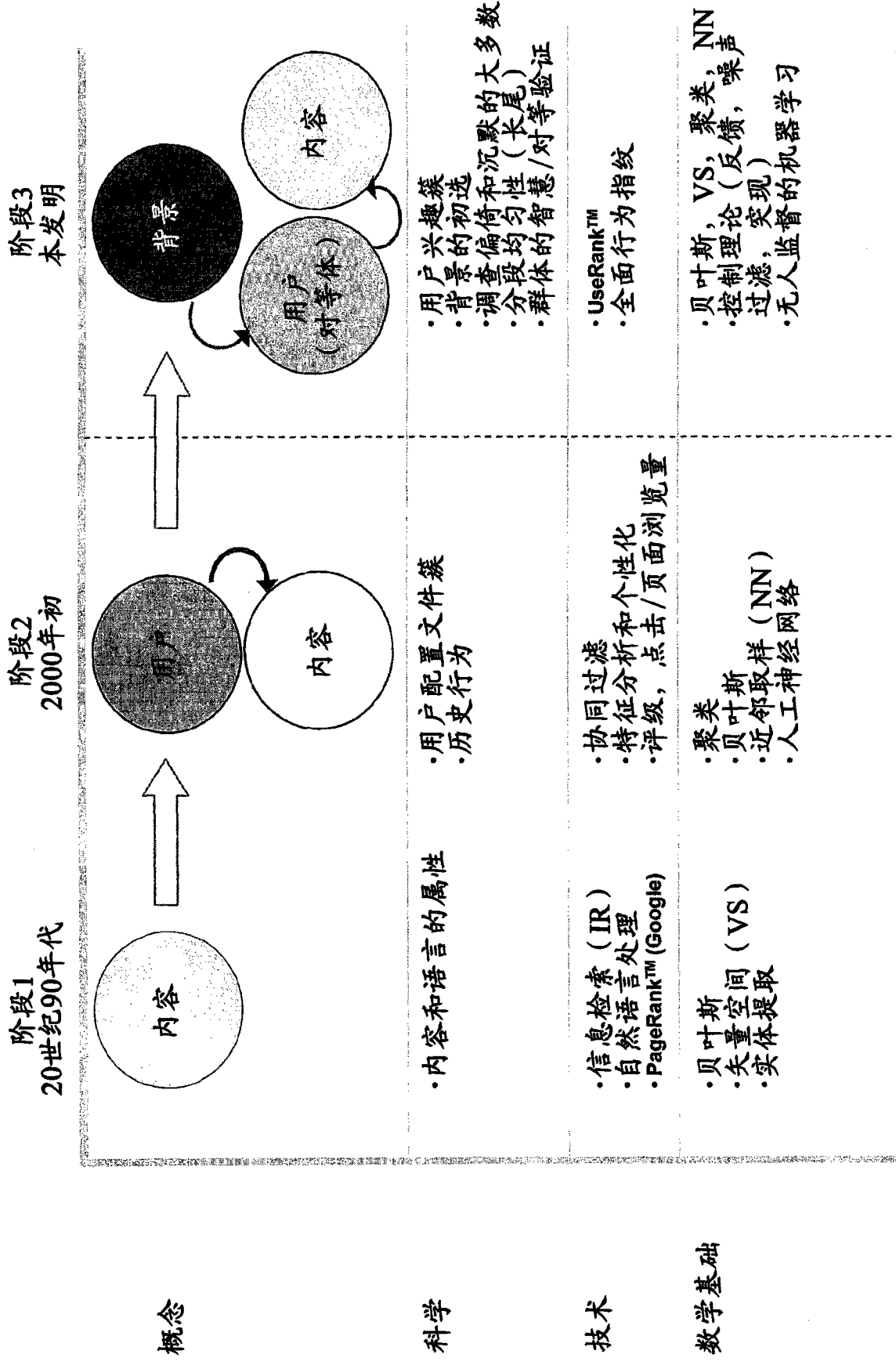
[0184] 代码的结构被分解,以使得共同的代码被使用。共同的代码或一段代码被用于所有类型的推荐。因此,标记可能要求共同的代码。政策可能是顾客所特有的,或它可能是用户所特有的。这涉及怎样做出推荐。政策信息被分别加载。然后,具体的标记所需的代码也被加载。例如,如果有一个用于最流行的推荐的标记,并且另外一个推荐,那么这些标记被加载。然后系统加载执行特定的标记操作所必需的代码。另一种类型的标记或代码是观察用户做什么的观察者,如上所述。用户看不见标记,但它仍然是标记。标记必须也能够与系统进行通信。这些功能的每一个都由一段代码如 JavaScript 执行,被加载到网页中。JavaScript 通过动态脚本注入被加载到页面。

[0185] 本发明的此方面认识到网页可变得更加动态。例如,存在诸如 HTML 和 JavaScript 的事物,导致事物被动态添加到页面。如果脚本标记需要一段特定的信息,那么那个代码被注入该页面。例如,如果需要政策,那么 JavaScript 标记被注入到访问该政策的页面。该

政策可要求信息的进一步注入,例如处理机。因此,处理机现在作为 JavaScript 段被动态地注入页面。以相似的情况处理通信。

[0186] 上述方法的一个优势是提供使用第一方 cookies 或第三方 cookies 的选择。顾客访问的主要网页可能是商人的网站。该商人与系统通信以得到关于推荐等的信息。在此情况下,该商人是第一方,系统是第三方。作为动态脚本注入的结果,本发明允许 cookies 设置在第一方域和第三方域中。设置第一和第三方 cookies 的能力还允许本发明跨域运行。

[0187] 虽然这里描述的发明是参考优选实施方式的,但是本领域的技术人员很容易领会,在不偏离本发明的精神和范围的情况下,本文所阐述的这些可由其他应用代替。因此,本发明应只受限于下面所包括的权利要求。



概念

科学

技术

数学基础

图 1

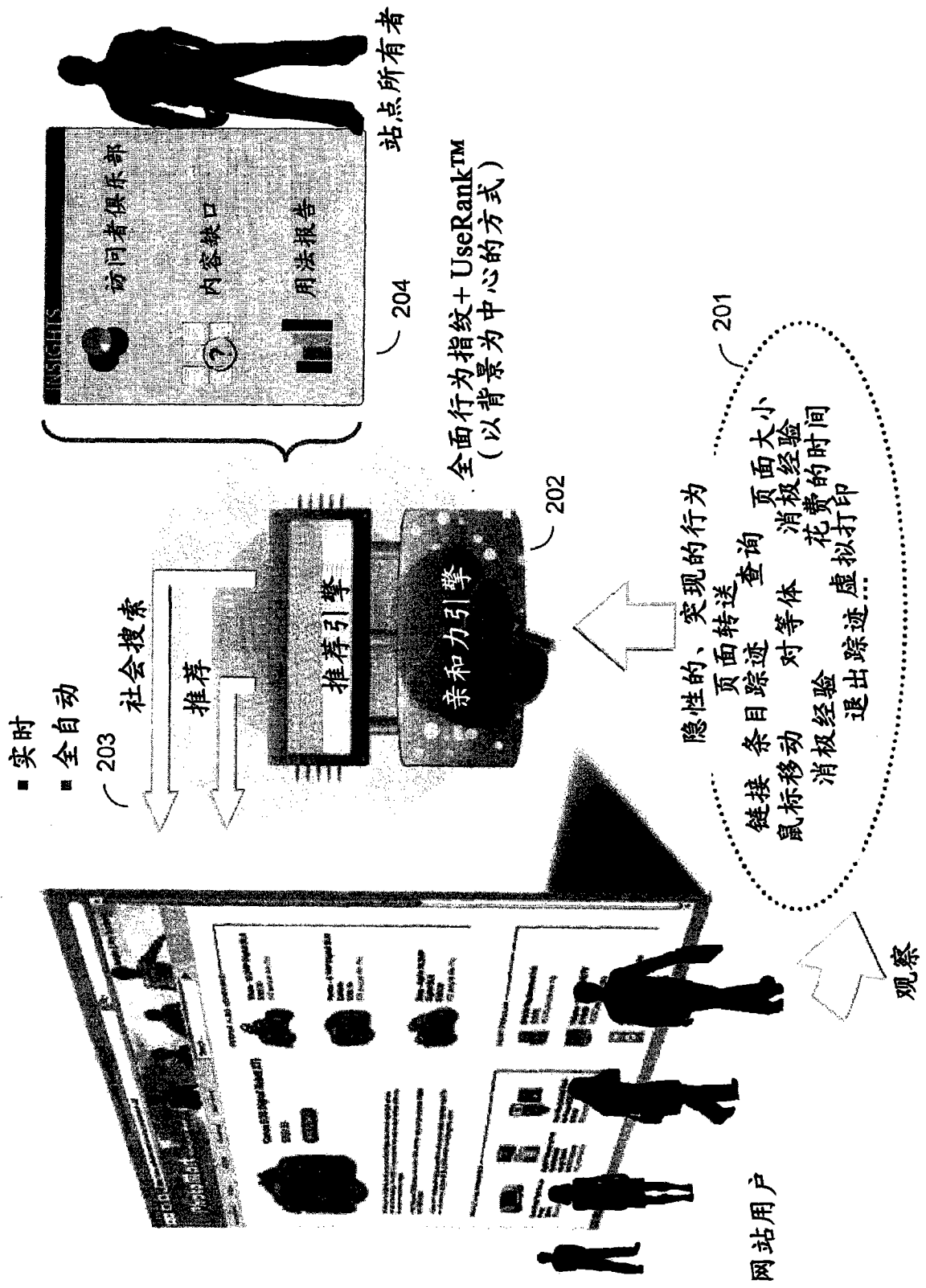


图 2

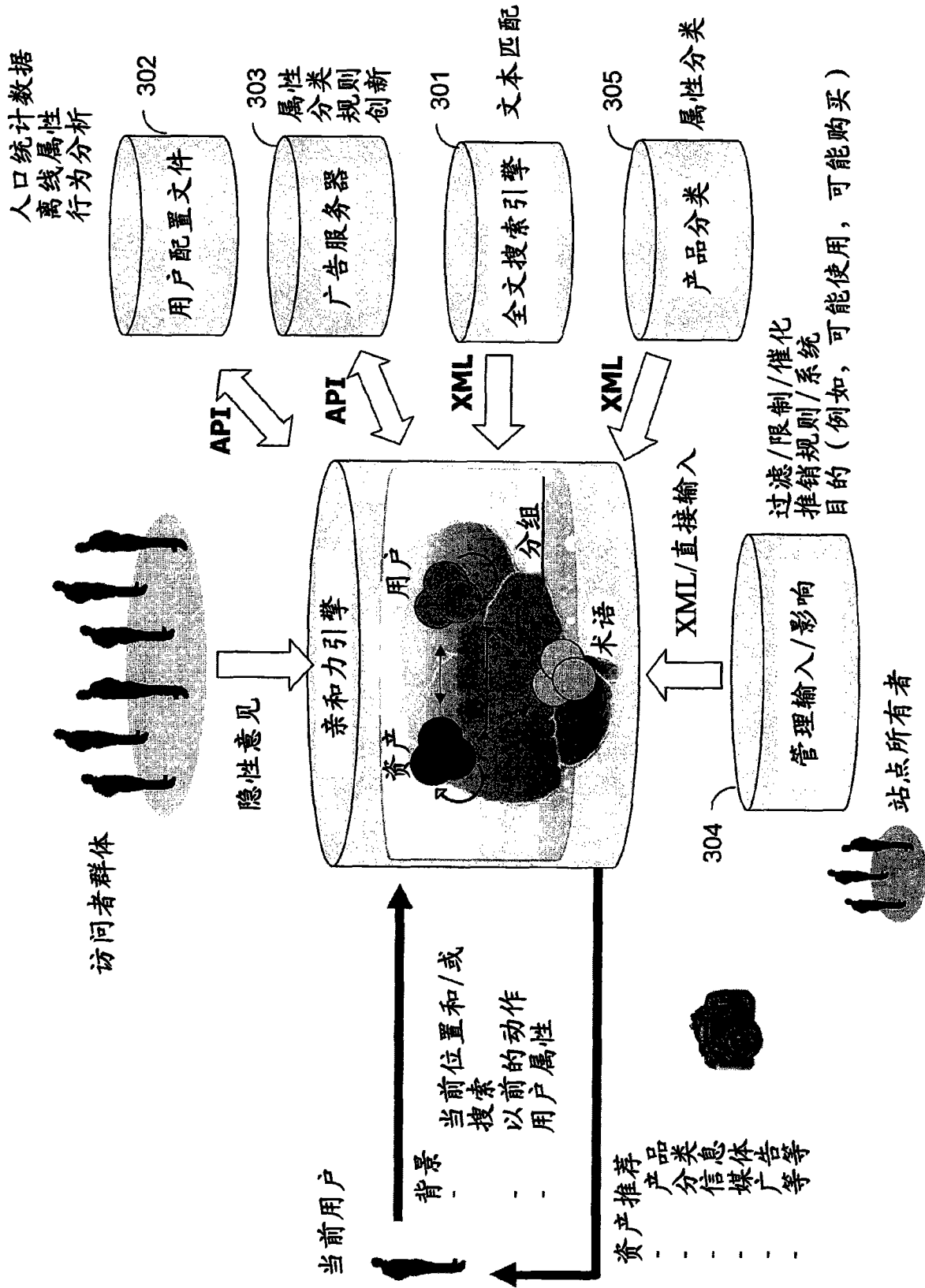


图 3

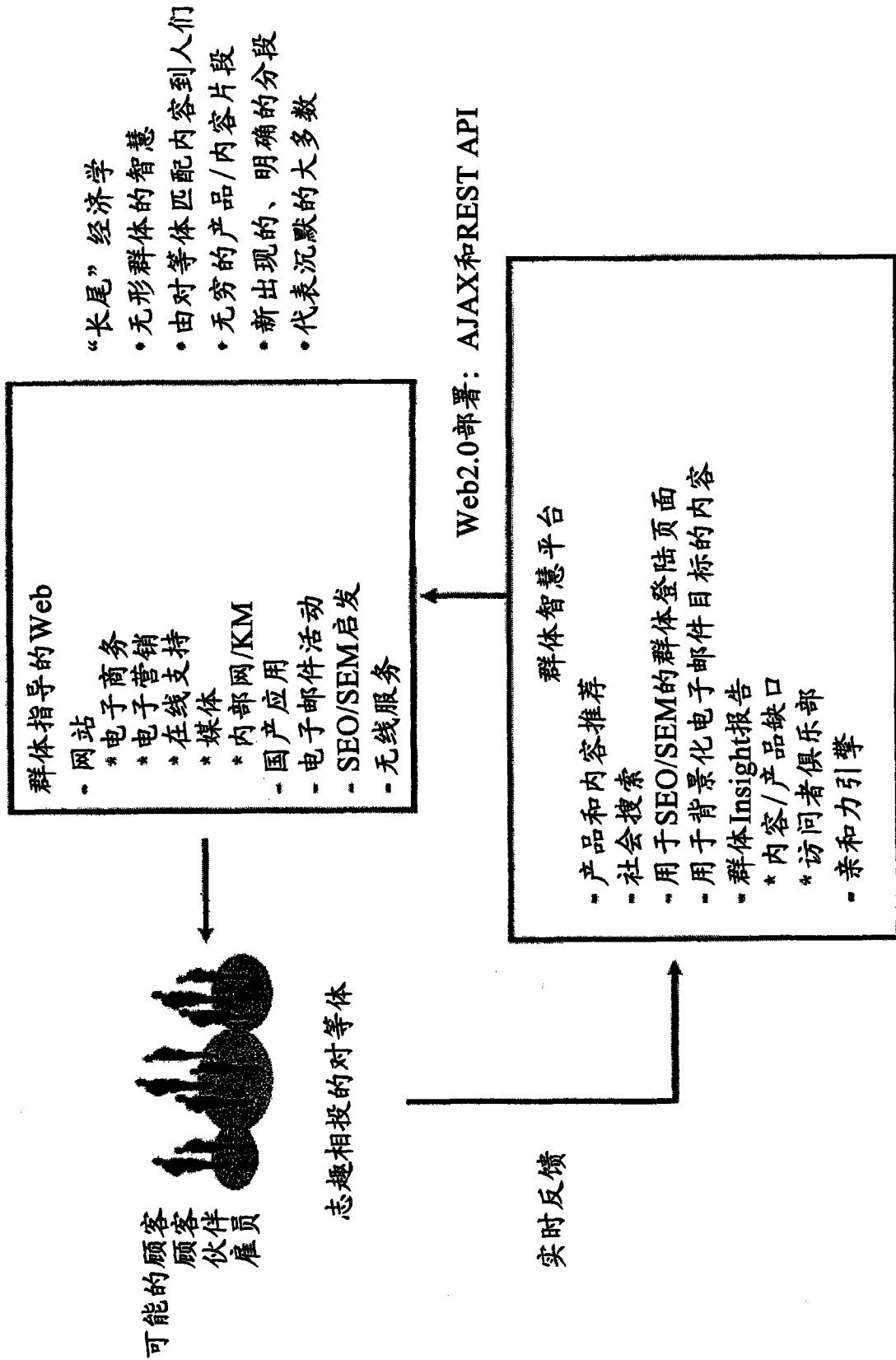


图 4





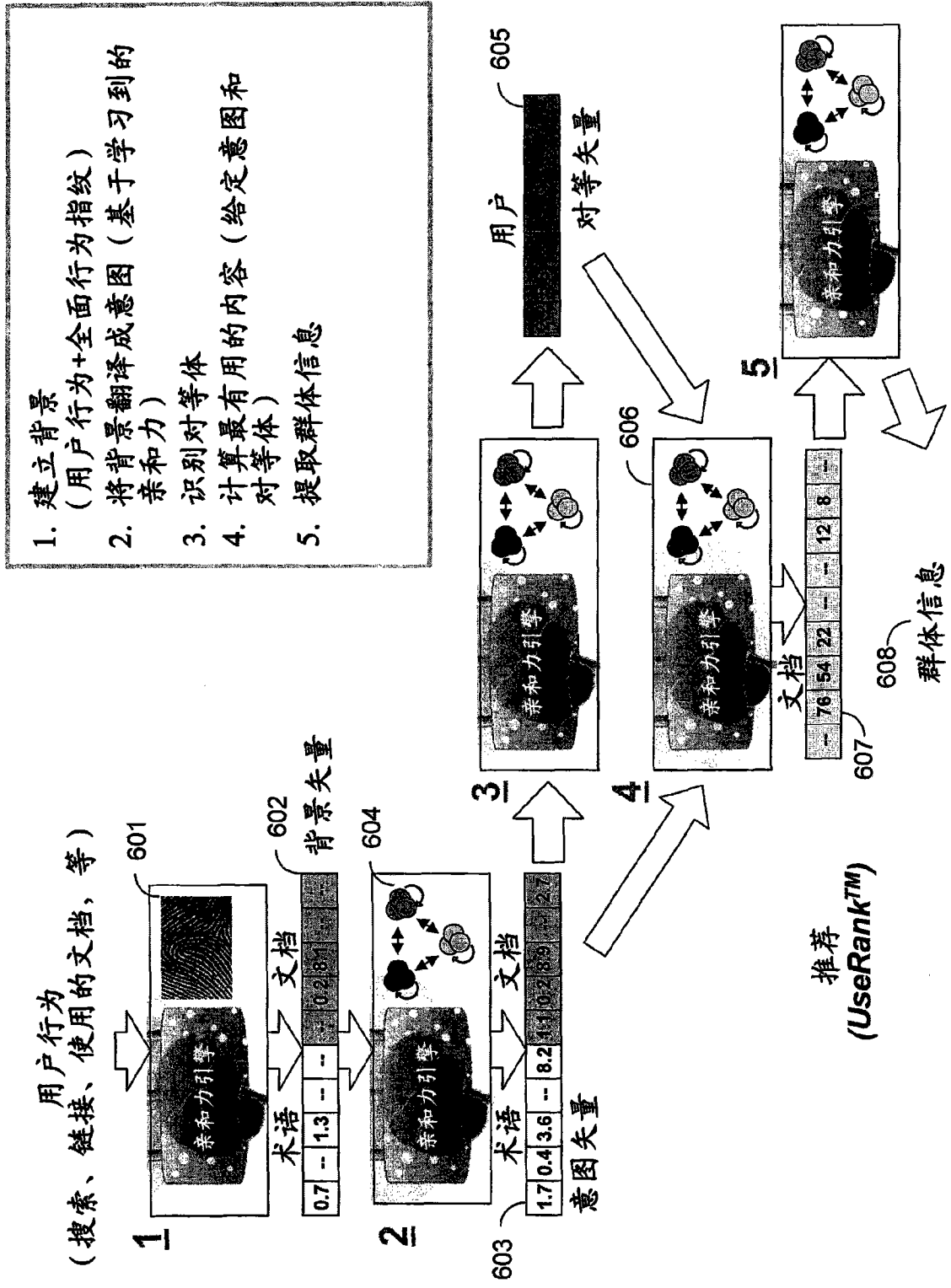
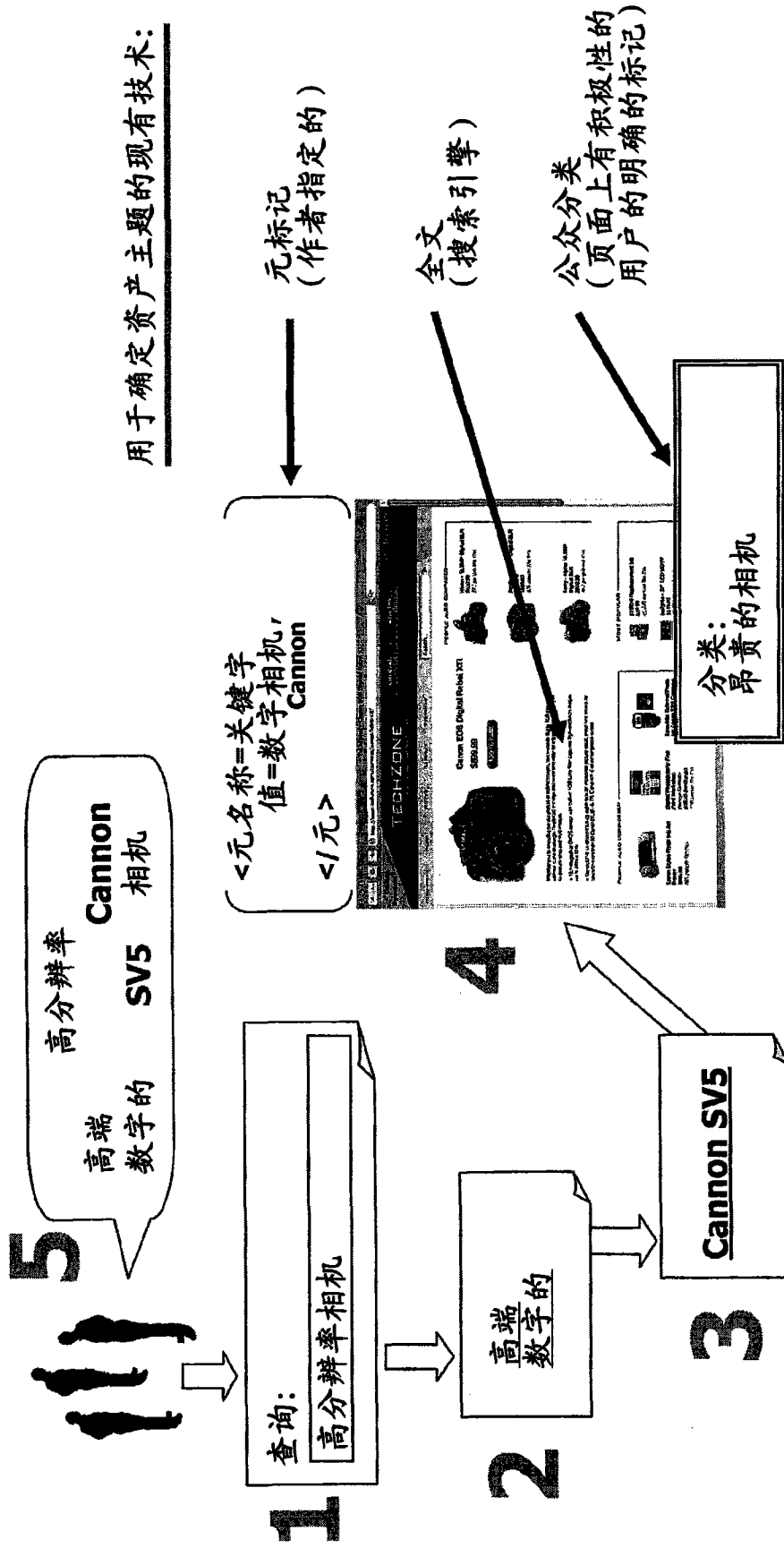


图 6



- 1-3: 用法背景—搜索查询、使用的链接文本、近来的页面, 等
- +4: 资产用法
- =5: 虚拟公众分类---通过隐性的用法/证实, 由整个群体所搜集的与资产相关的术语

图 7

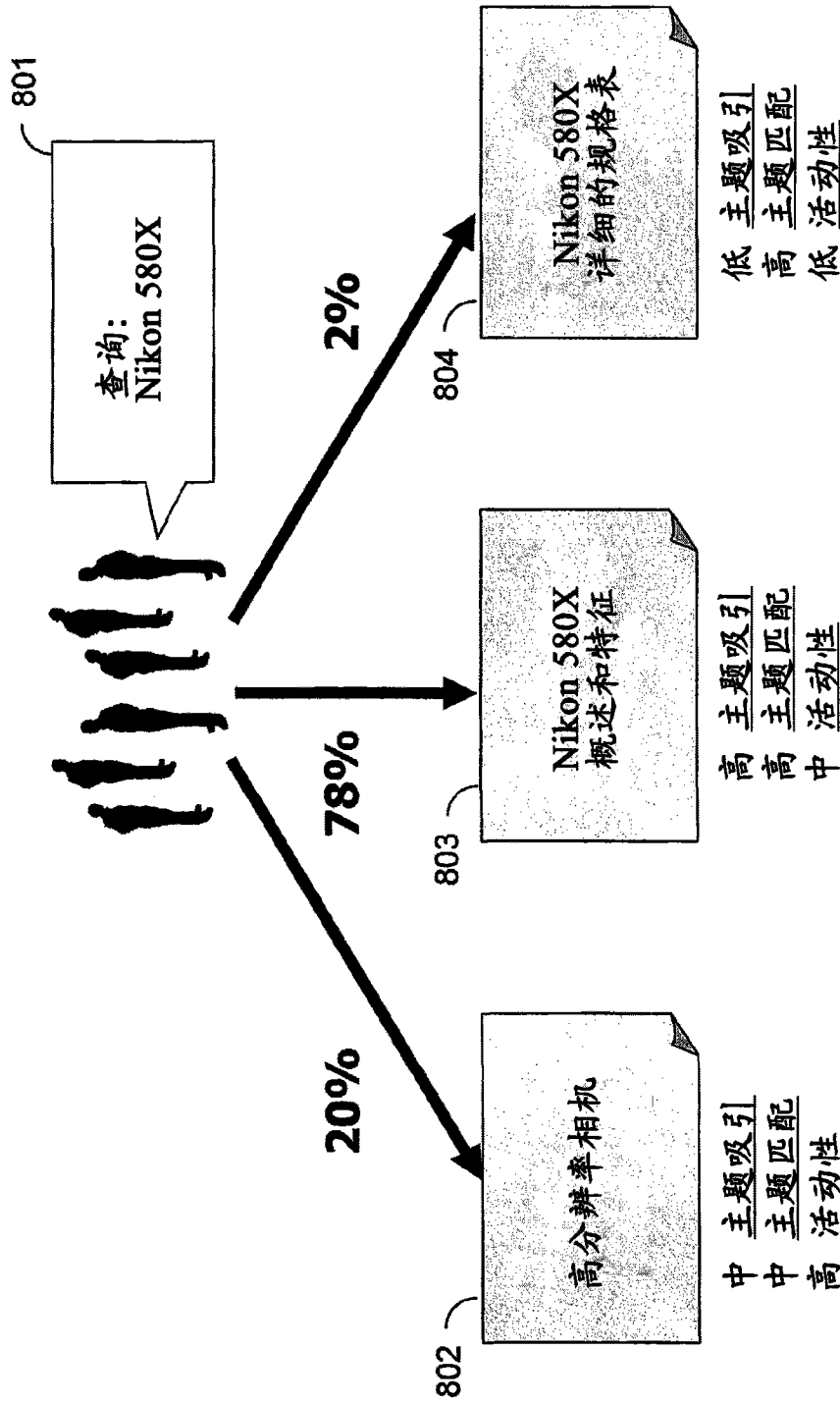


图 8

主题吸引: 在寻找“Nikon 580X”的人们中, 有多少比例在此资产中找到价值。即, 此资产有多吸引“Nikon 580X”的搜索者。

主题匹配: 在此文档中找到价值的人们, 有多少比例正在寻找“Nikon 580X”。即, 此资产对于主题“Nikon 580X”的关注程度如何。

活动性: 多少人已经在此资产中找到价值 (以任何原因)

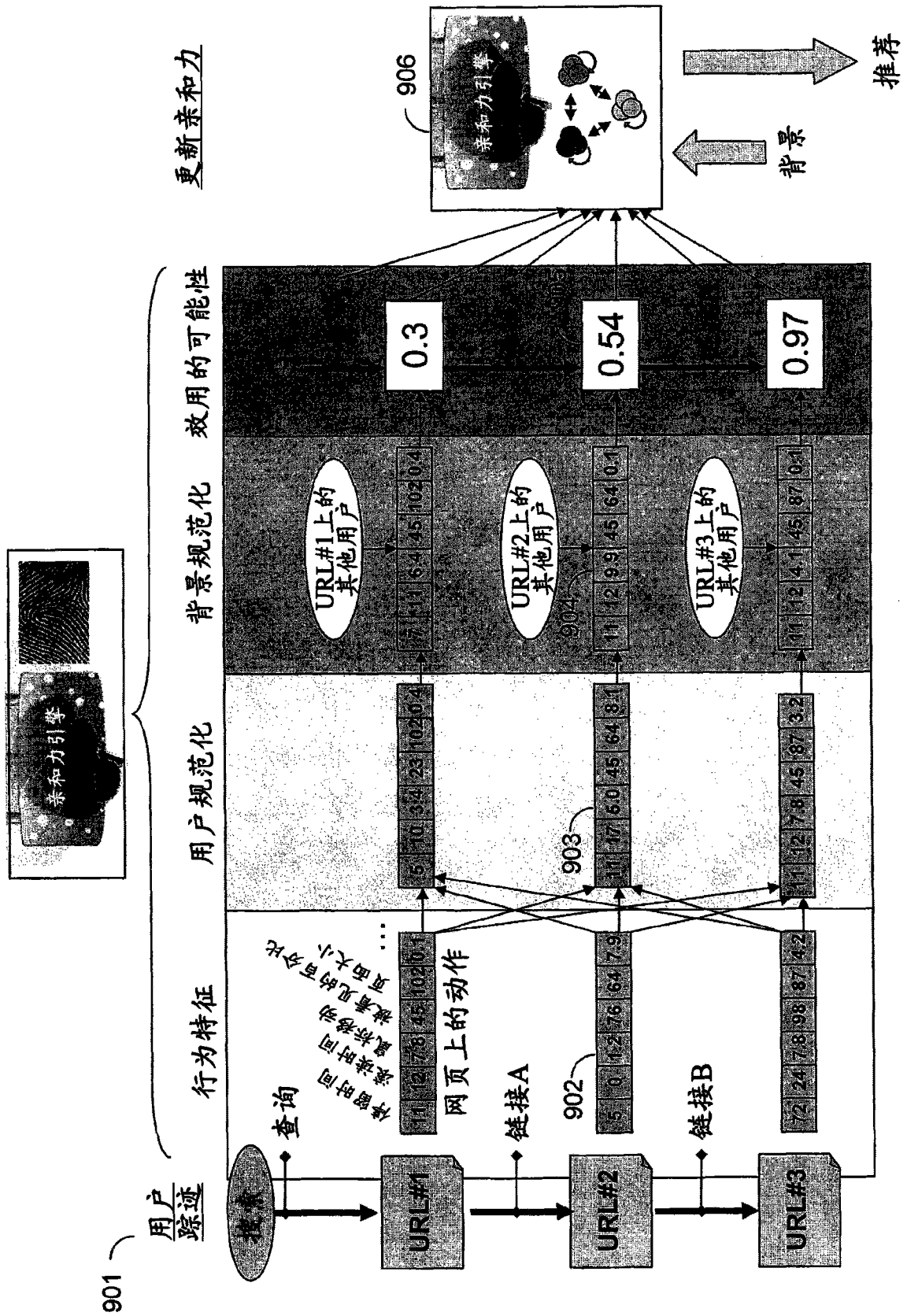


图 9

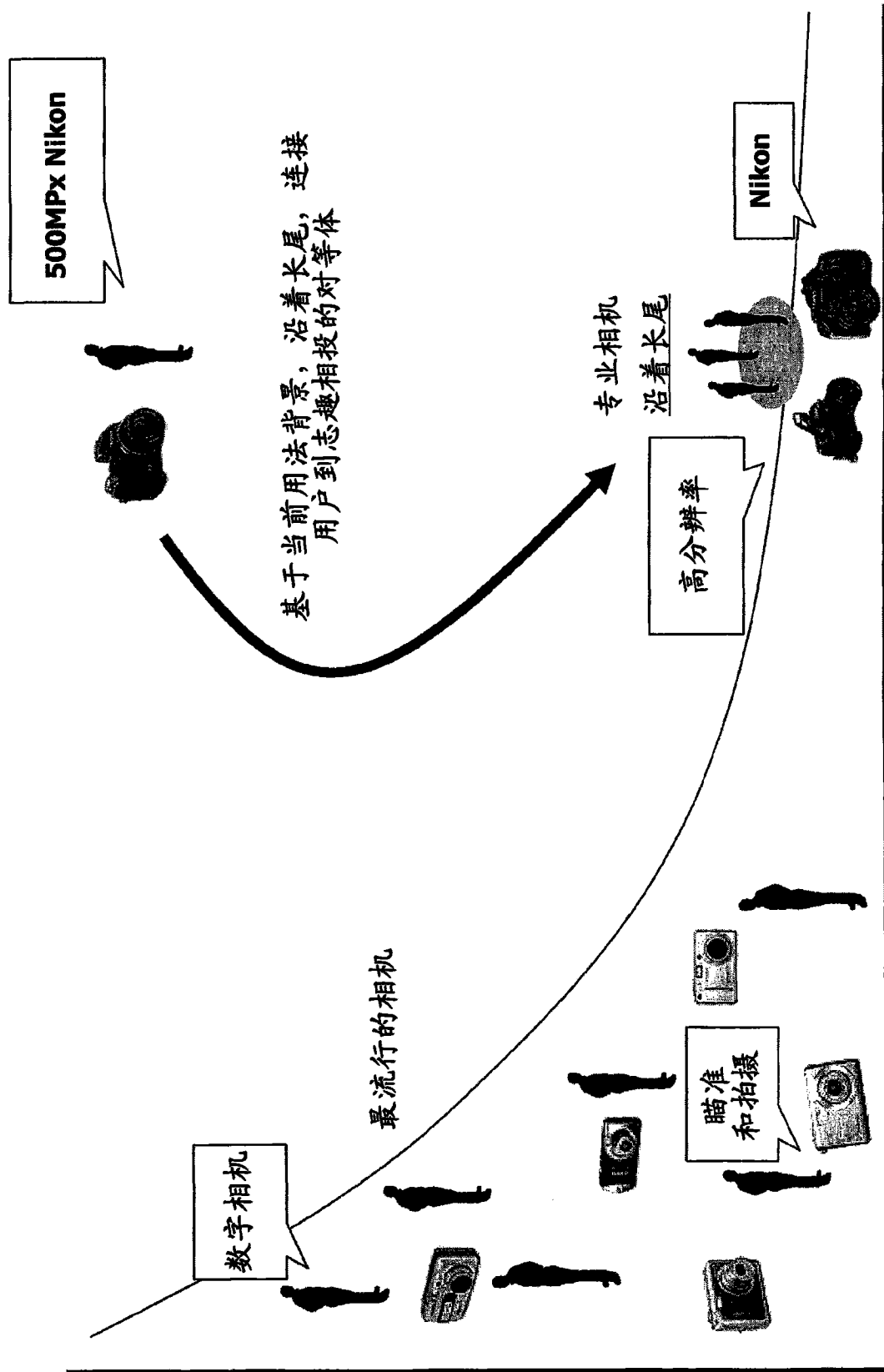


图 10

$$P(X = x' | C = c) = \text{sim}_X(x', \frac{1}{U} \sum_{u=1}^U x_u * \text{sim}_C(c_u, c))$$

X: 兴趣的变量  
 x': 变量的候选值  
 sim<sub>x</sub>: 比较X值的相似性函数  
 C: 背景  
 c': 当前背景  
 sim<sub>c</sub>: 比较背景的相似性函数  
 U: 用户  
 x<sub>u</sub>: 特定的用户的X值 (估计)  
 c<sub>u</sub>: 用户的估计的背景

▪ 例如，给定特定的背景c' (诸如导航/搜索历史)，确定特定的文档x'是最好的推荐的可能性

- 对所有的用户文档估计执行“加权平均”，其中权重是基于在估值时那个用户的背景和当前背景之间的相似性。文档被在一个选定的属性例如术语-空间代表上平均。(注意: WOC等式也可被用来得出术语空间代表)

- 带入此平均的“群体智慧”值并使用适当的相似性矩阵来将它与候选文档比较，得出该候选是“最好”的可能性

▪ 例如，如果当前背景和所有的估值背景相同，那么此等式简化为简单的平均。Surowiecki的“猜测奶牛的情况”

图 11

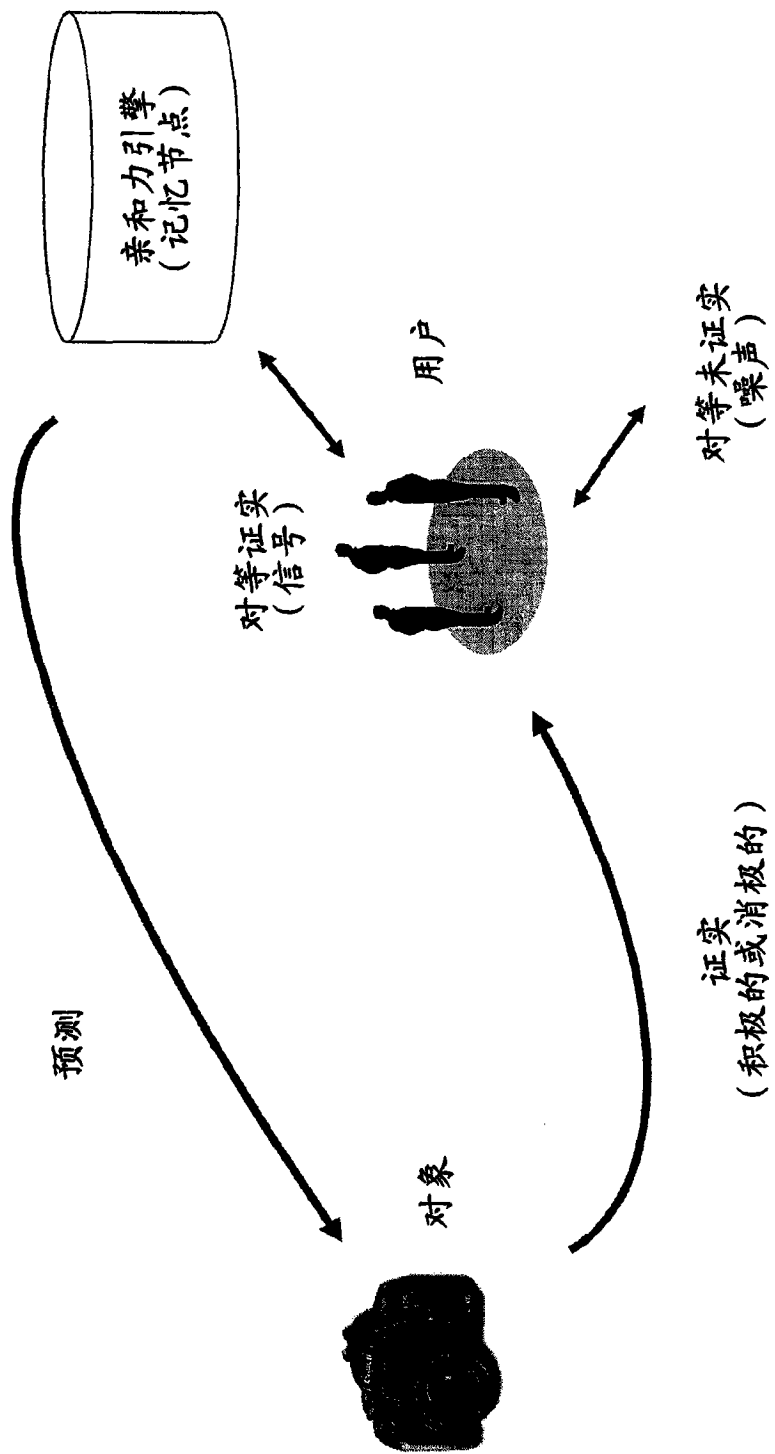


图 12

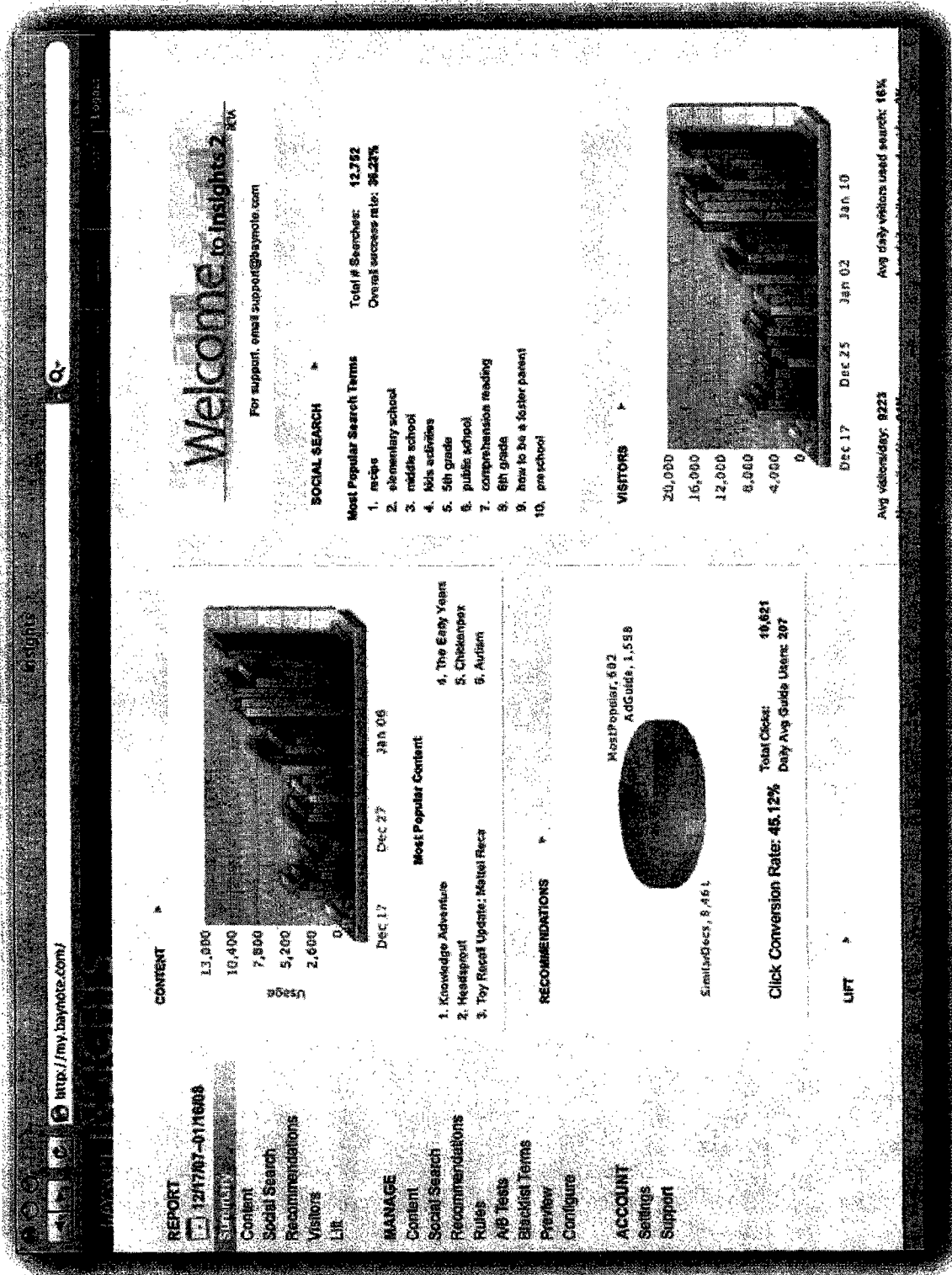


图 13



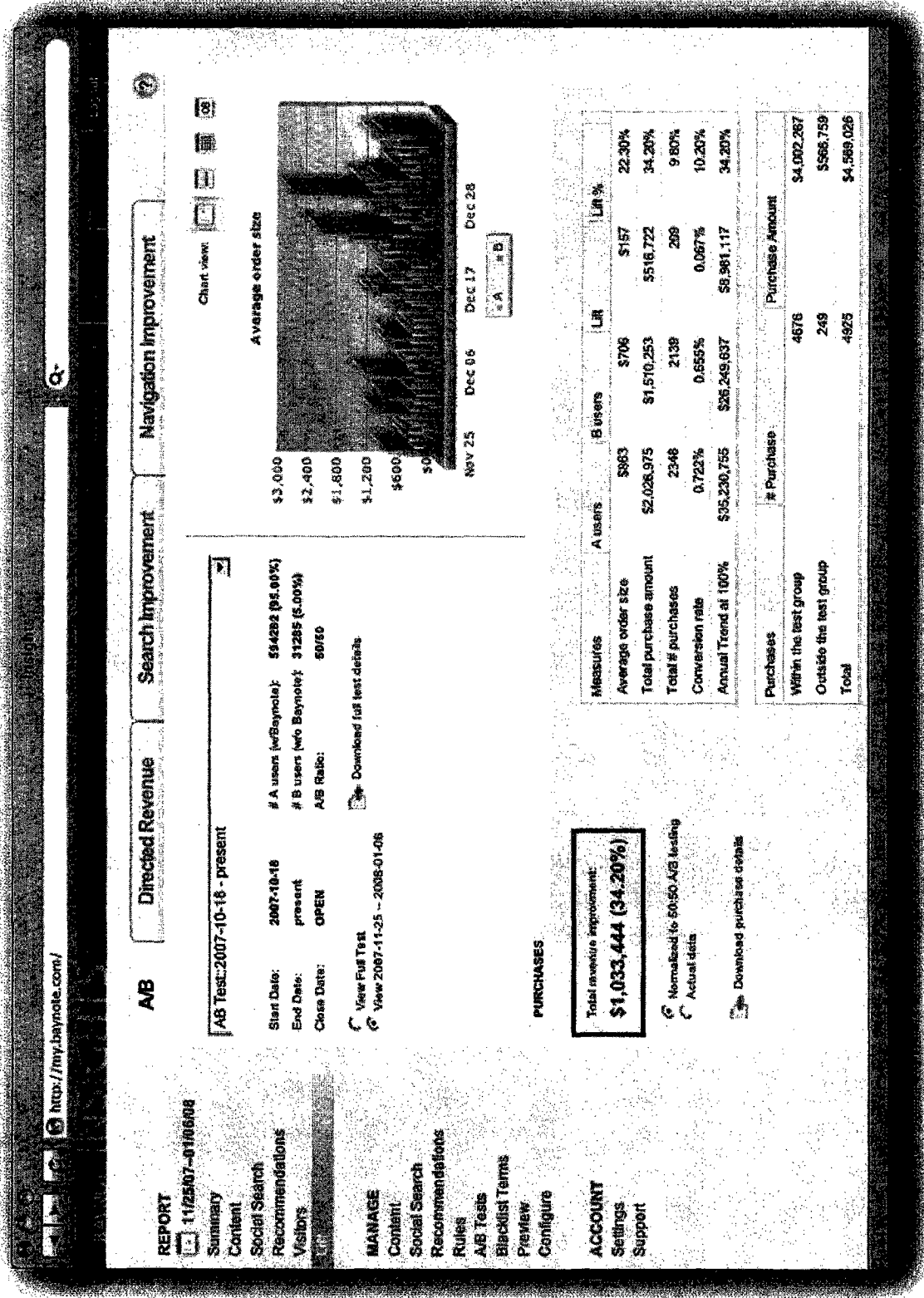


图 14

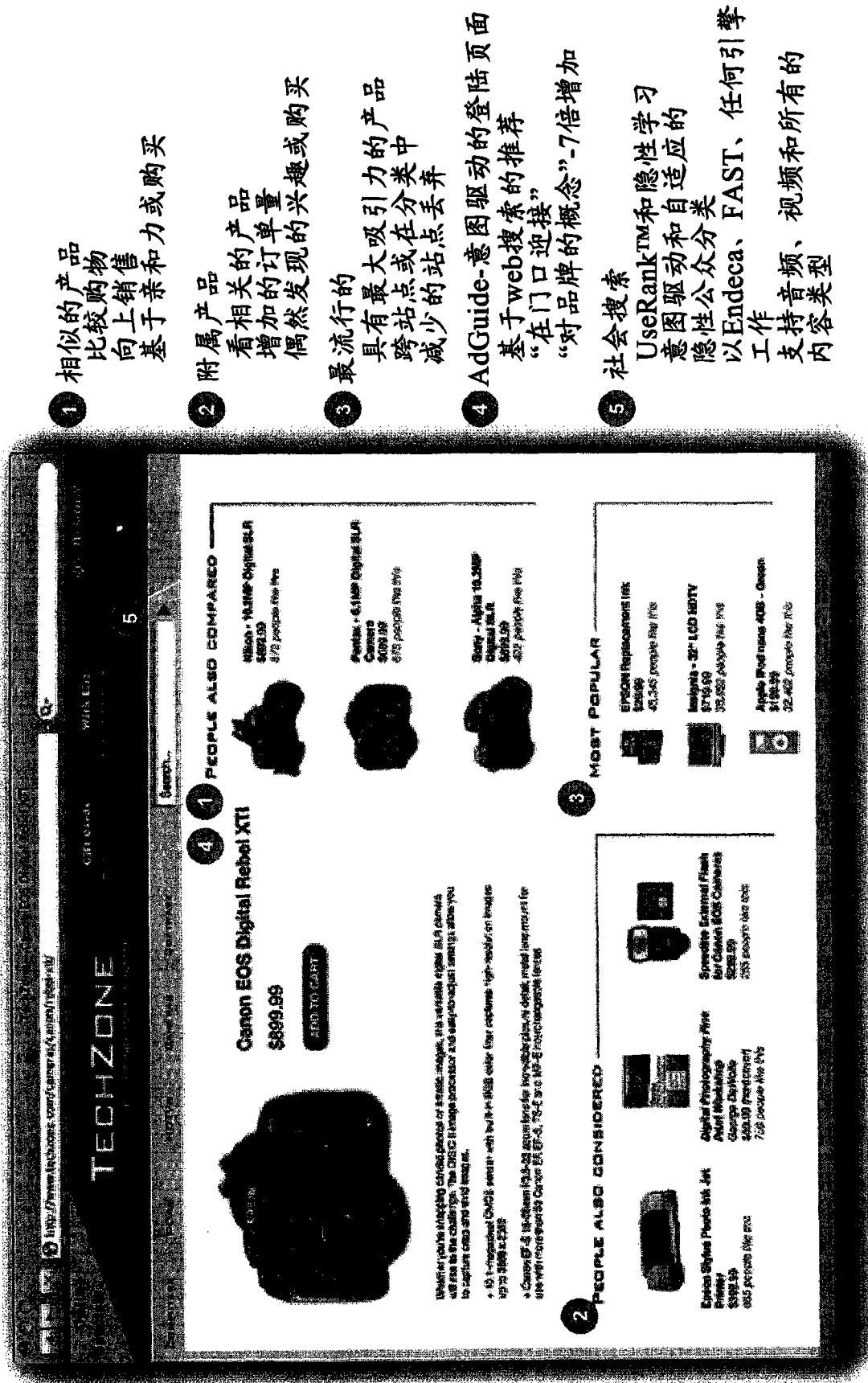


图 15

home | flights | hotels | cars | vacation packages | cruises | activities | deals & destinations | maps | corporate travel

Welcome - Affiliates/Member? Sign in

**1-800-551-2534**

**Start search over**

Change your search

Check-in: (mm/dd/yyyy)  
mm/dd/yyyy

Check-out: (mm/dd/yyyy)  
mm/dd/yyyy

Hotel name: \_\_\_\_\_

Hotel class: \_\_\_\_\_

Show all

Hotel type: \_\_\_\_\_

Show all

**Go**

**Traveler Recommends**

The Alex \*\*\*\*  
East Midtown / New York  
Hotel info: 800-246-3357

The Bryant Park Hotel \*\*\*\*  
Bryant Park-Adams / New York  
Hotel info: 800-246-3357

Sofitel New York \*\*\*\*  
Times Square/Theater District / New York  
Hotel info: 800-246-3357

Gramercy Park Hotel \*\*\*\*  
Union Square - 4 Blocks / New York  
Hotel info: 800-246-3357

Omni Berkshire Place \*\*\*\*  
Broadway/6 Blocks / New York  
Hotel info: 800-246-3357

The Ritz-Carlton \*\*\*\*  
Fifth Avenue/13 Blocks / New York  
Hotel info: 800-246-3357

**Lodging in New York City (and vicinity)**

Type your dates here, then click "Go" to check availability

Booking online is a snap - or call 1 (800) 509-1949 to book by phone.

Hotel list view | Area map view | How's map view

Show hotels in this area:  
New York City (and vicinity) (All areas) **Go**

Not what you're looking for? Choose a different destination

Page 1 of 14

Expedia Picks | Hotel Name | City | Hotel Class

**Park Central New York Hotel**  
\*\*\*\* New York, NY  
Area: Midtown (Times Square Broadway)

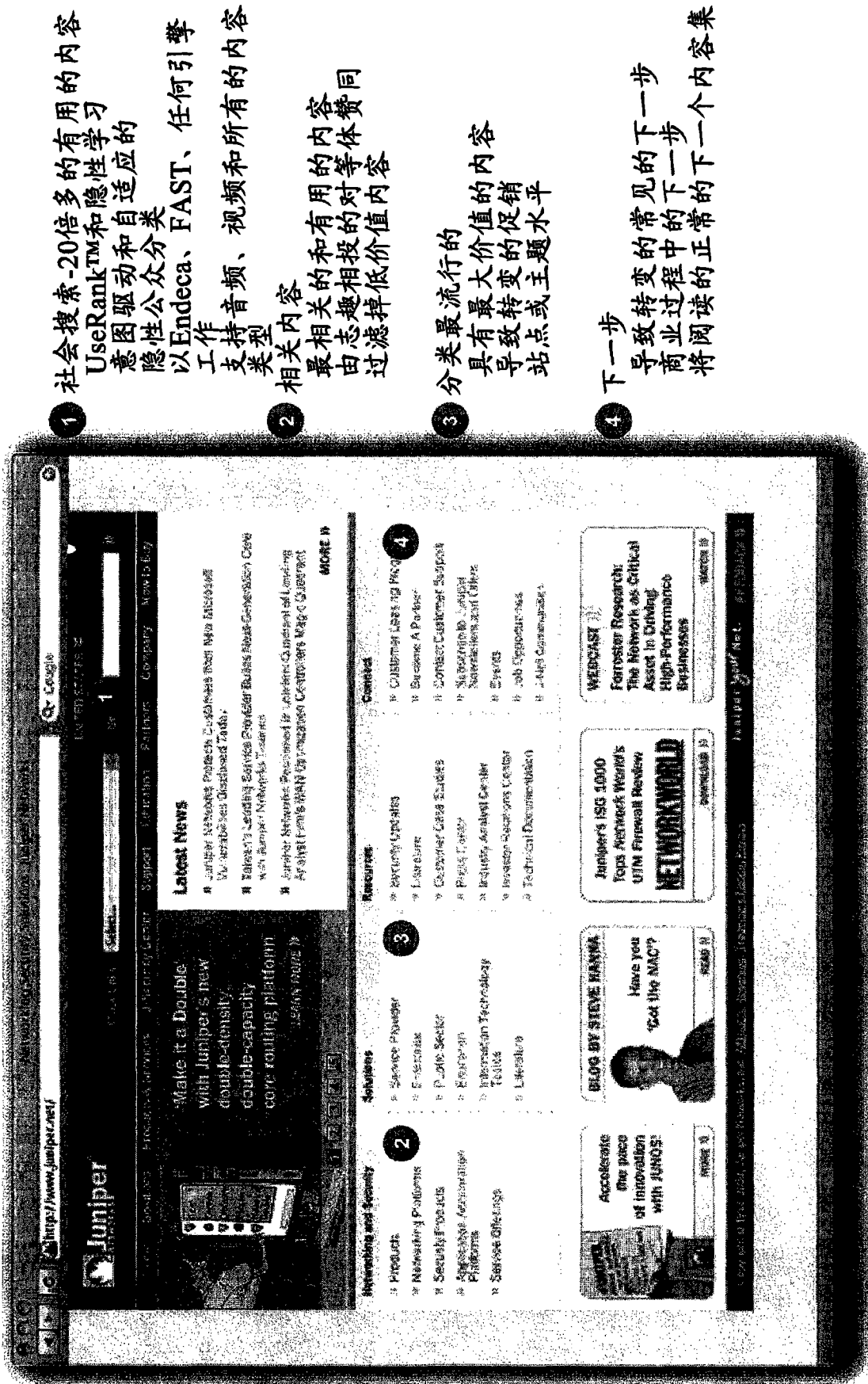
**Midtown hotel three blocks south of Central Park**  
This 25-story hotel offers a desirable location diagonally across from Carnegie Museum of Art, one block from Broadway, two blocks from Fifth Avenue shopping, and ...

More lodging info

Enjoy rates from \$219.00 per night between Mar-27-2007 and May-26-2007. **Available**

**Milford Plaza**  
\*\*\*\* New York, NY  
Area: Midtown (Times Square Broadway)  
**Casual, 1,302-room hotel in Theater District**  
This 28-story hotel is at the corner of West 45th Street and Eighth Avenue, four blocks

图 16



1 社会搜索-20倍多的有用的内容  
UseRank™和隐性学习  
意图驱动和自适应的  
隐性公众分类

2 以Endeca、FAST、任何引擎  
工作支持音频、视频和所有的内容  
类型

3 相关内容的  
最相关的和有用的内容  
由志趣相投的对等体  
过滤掉低价值内容

4 分类最流行的  
具有最大价值的内容  
导致转变的促销  
站点或主题水平

5 下一步  
导致转变的常见的下一步  
商业过程中的正常的下一个内容集  
将阅读的

图 17

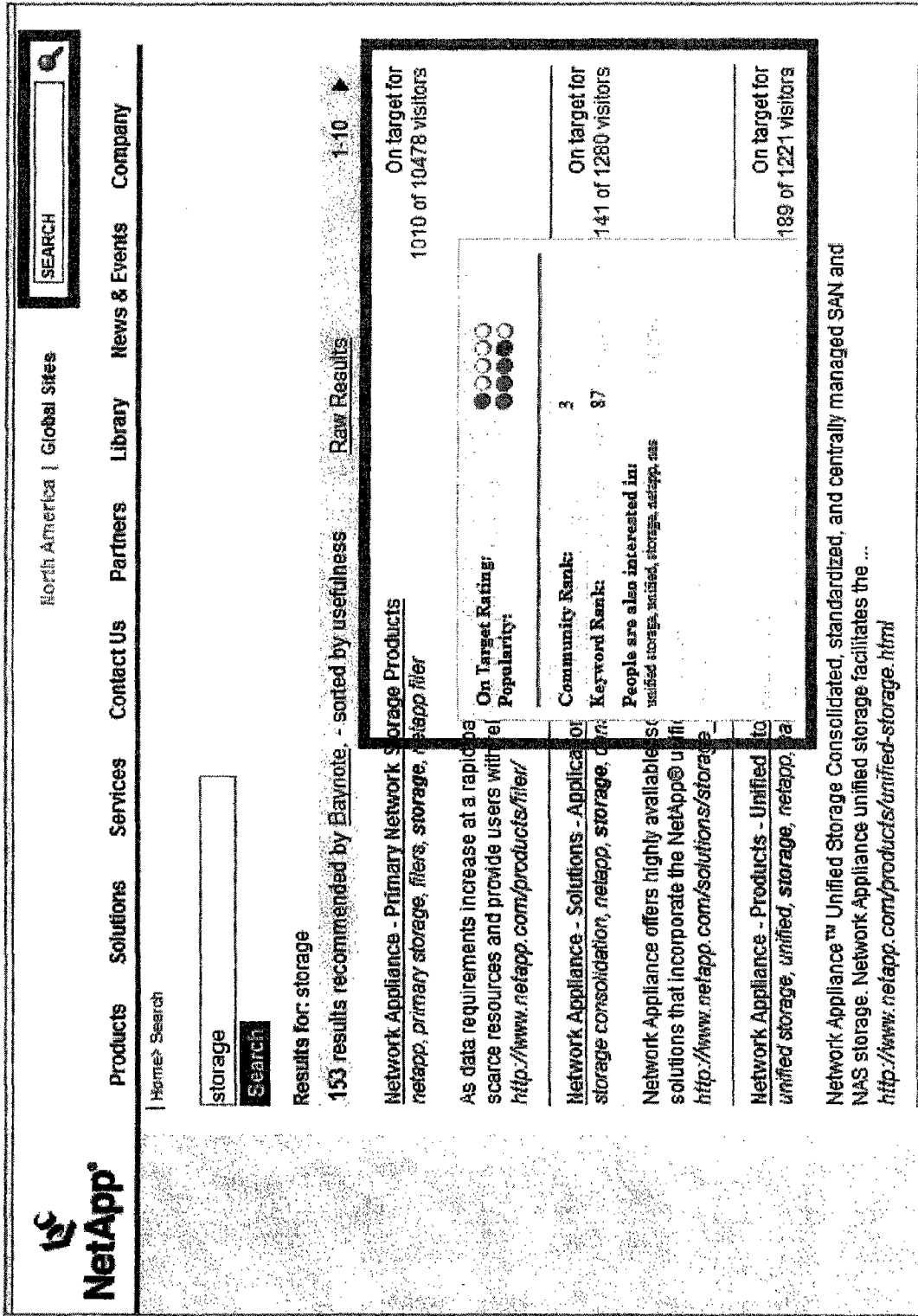
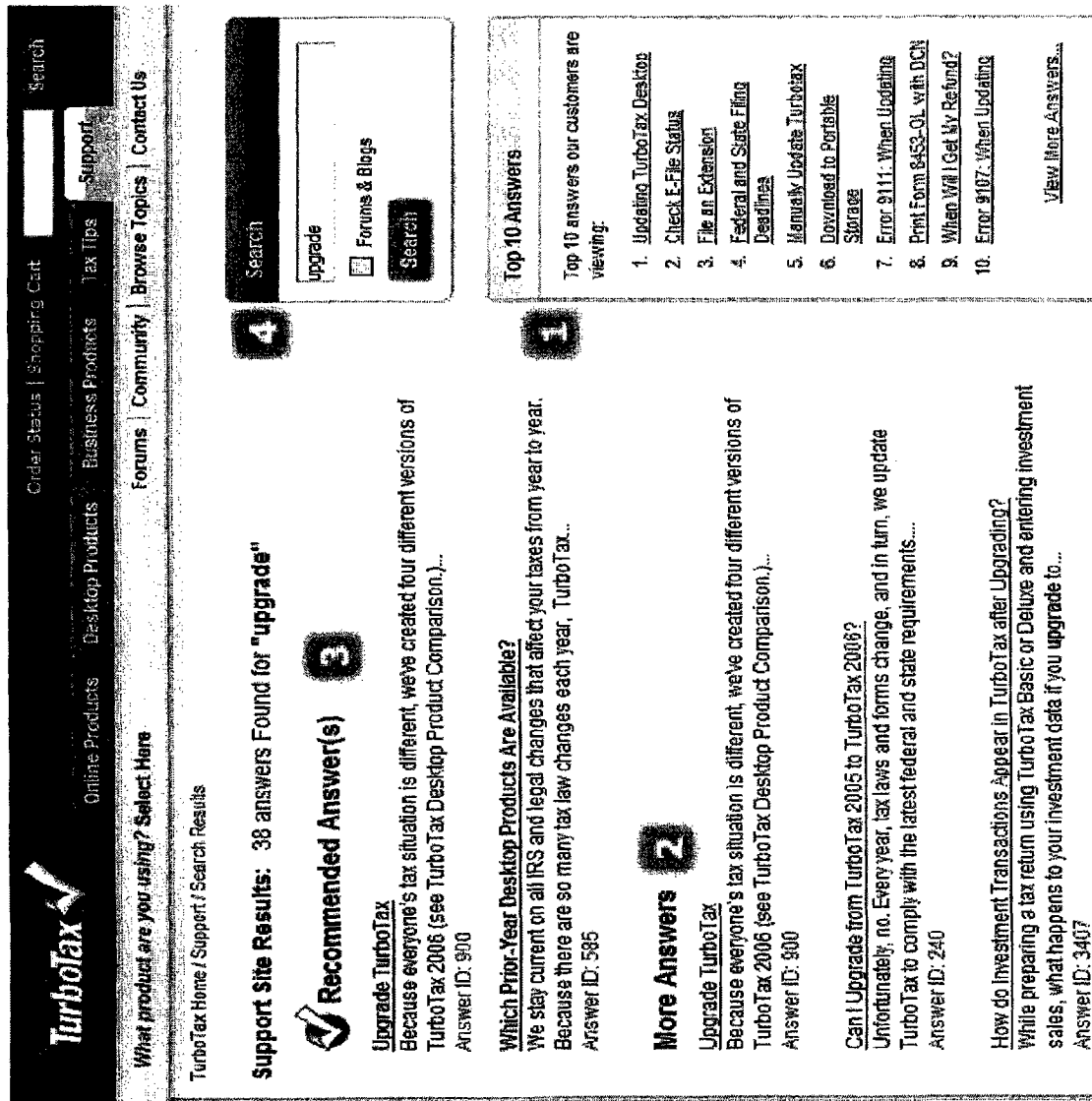


图 18



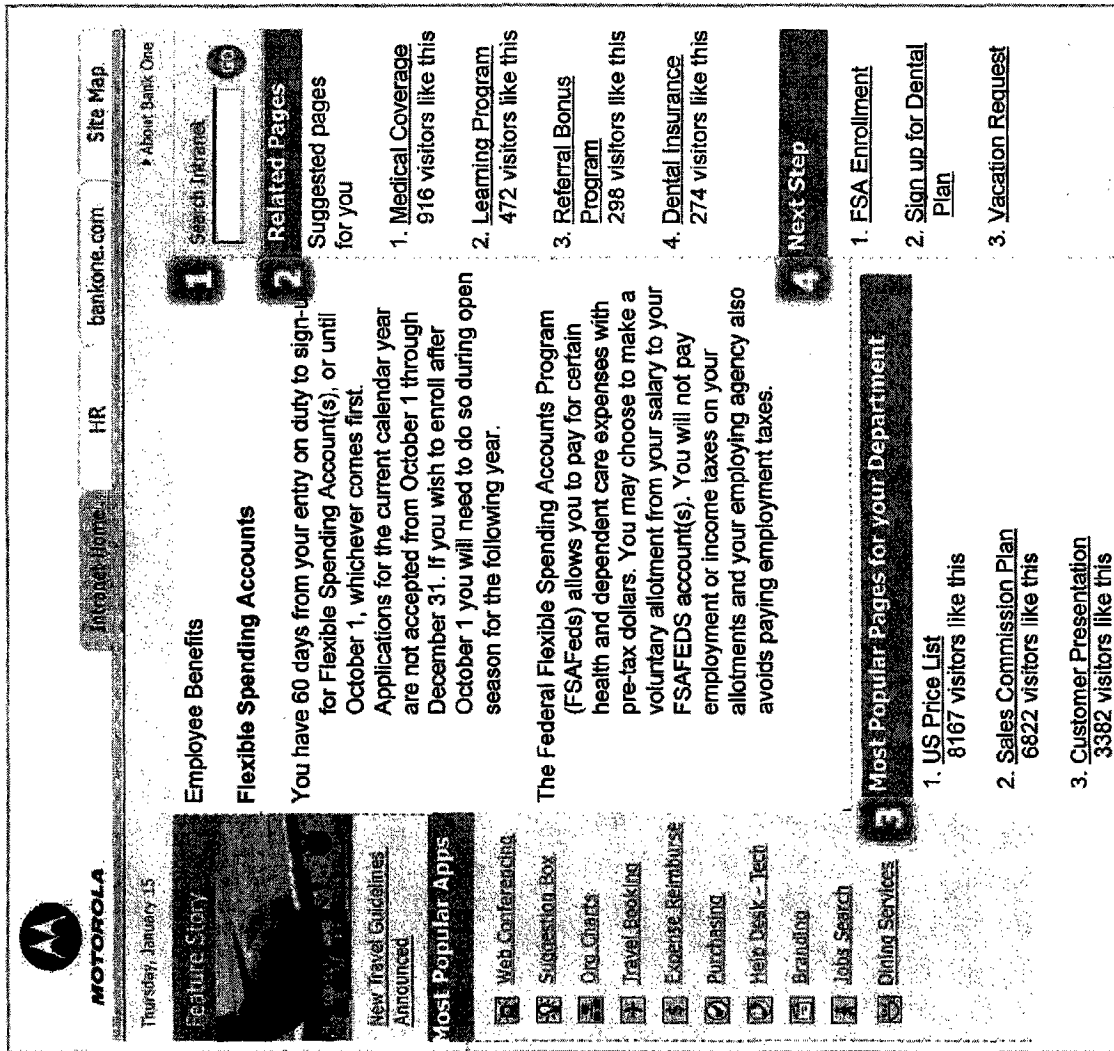
1 顶级问题解决方案  
最有帮助的文章  
最常见的问题  
帮助论坛和博客文章

2 相关帮助  
关于该题目的有深度的KB文章  
可选择的解决方案  
友好提示

3 常见的随后步骤  
解决方案的正常的下一步  
常见的下载  
附加的支持信息/联系

4 社会搜索  
20倍多的有用的内容  
UseRank™和隐性学习  
意图驱动和自适应的  
隐性公众分类  
以Endeca、FAST、任何引擎  
工作支持音频、视频和所有的  
内容类型

图 19



**1** 社会搜索  
 20倍多的有用的内容  
 UseRank™和隐性学习  
 意图驱动和自适应的  
 隐性公众分类  
 以Endeca、FAST、任何引擎工作  
 支持音频、视频和所有的  
 内容类型

**2** 相关页面  
 关于该主题的相似页面  
 由过去的对等体验所证明  
 启动偶然发现  
 过滤掉低价值内容

**3** 最流行的  
 具有最大价值的内容  
 公司内的顶级应用  
 部门所特有的或跨公司的

**4** 随后的步骤  
 人们从这里去满足他们的目的  
 商业过程中的下一个内容  
 将阅读的下一个内容集

图 20

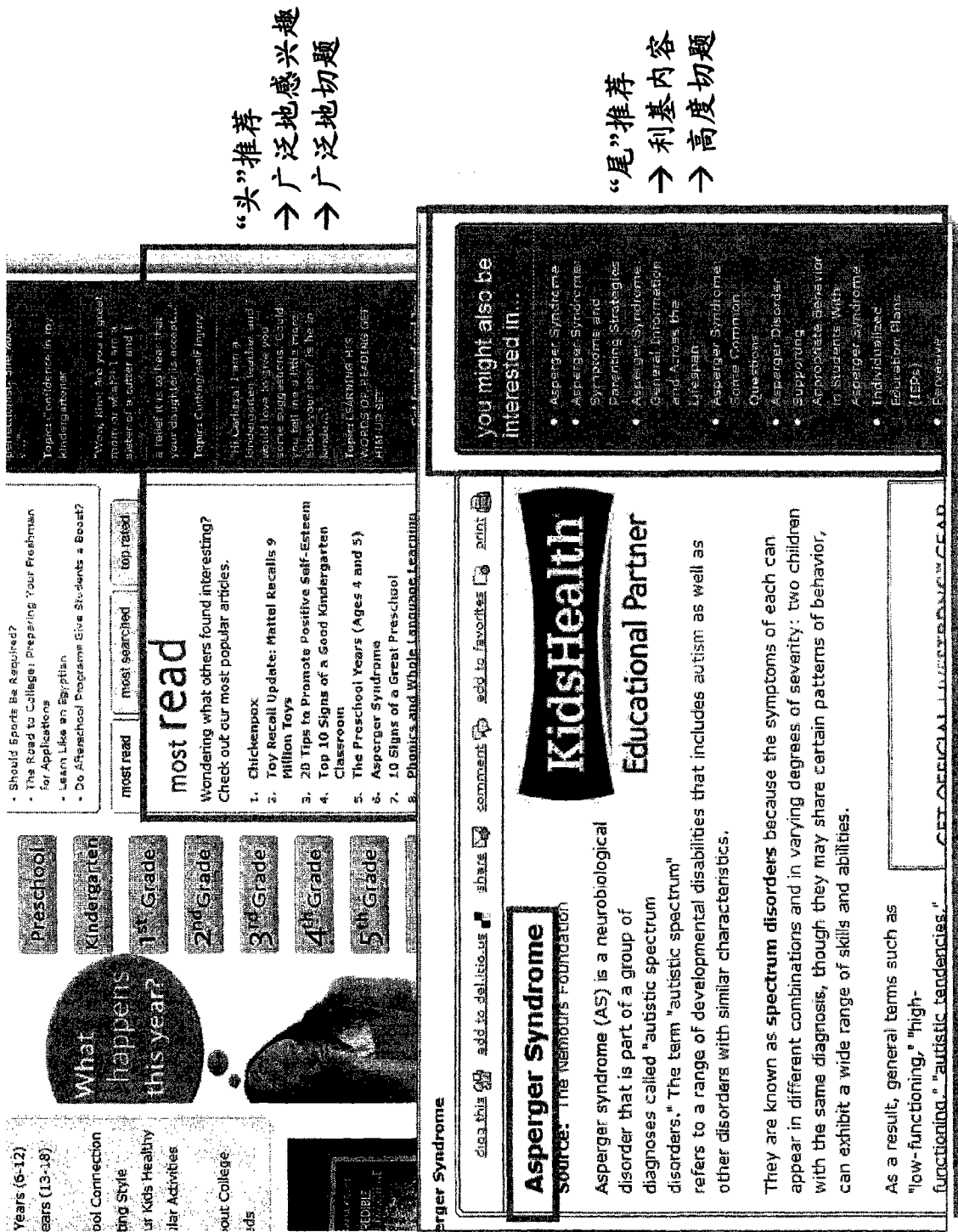
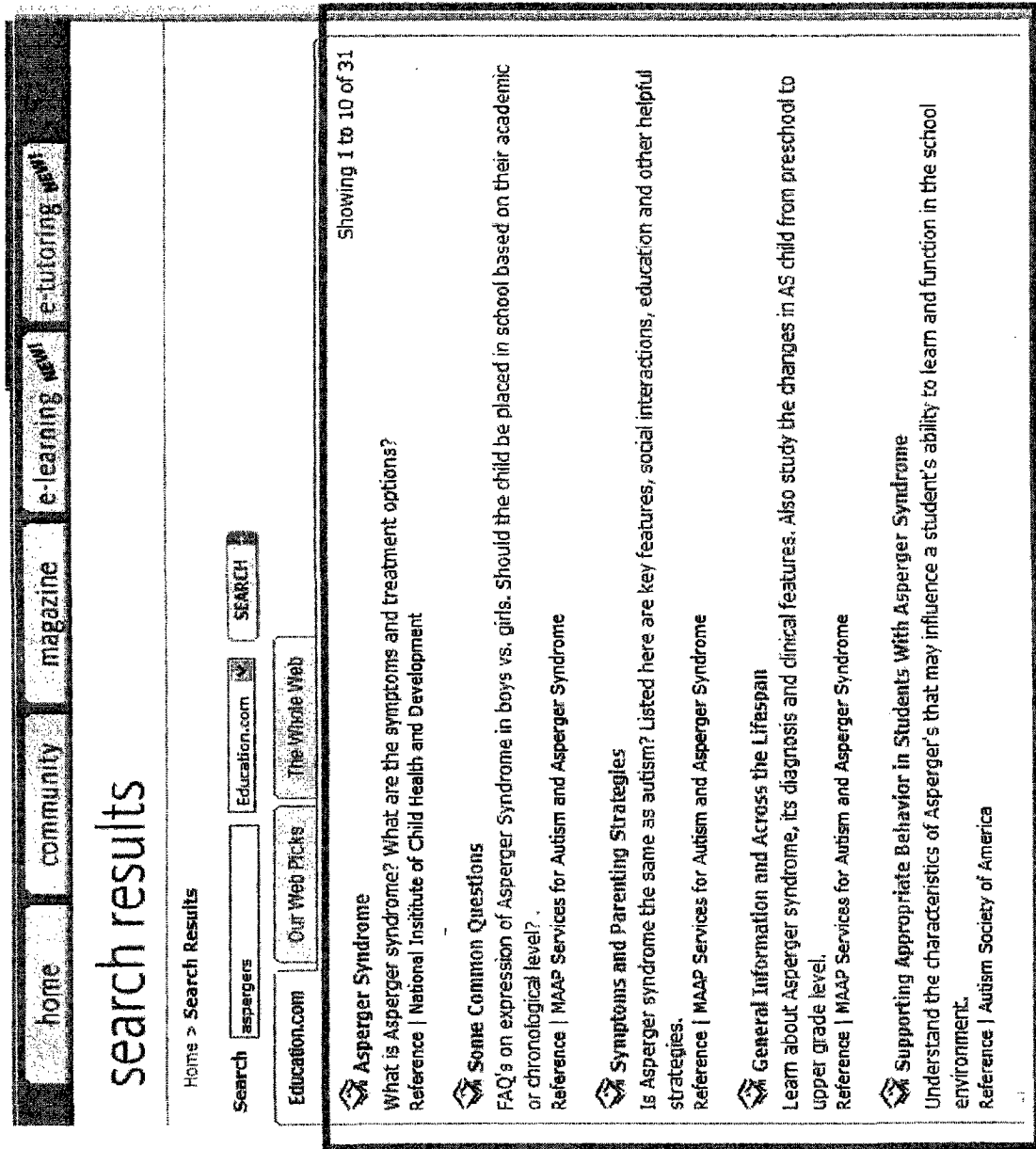


图 21





社会化驱动搜索

来自长尾搜索术语的利润

比关键字搜索10倍多的转变

图 22

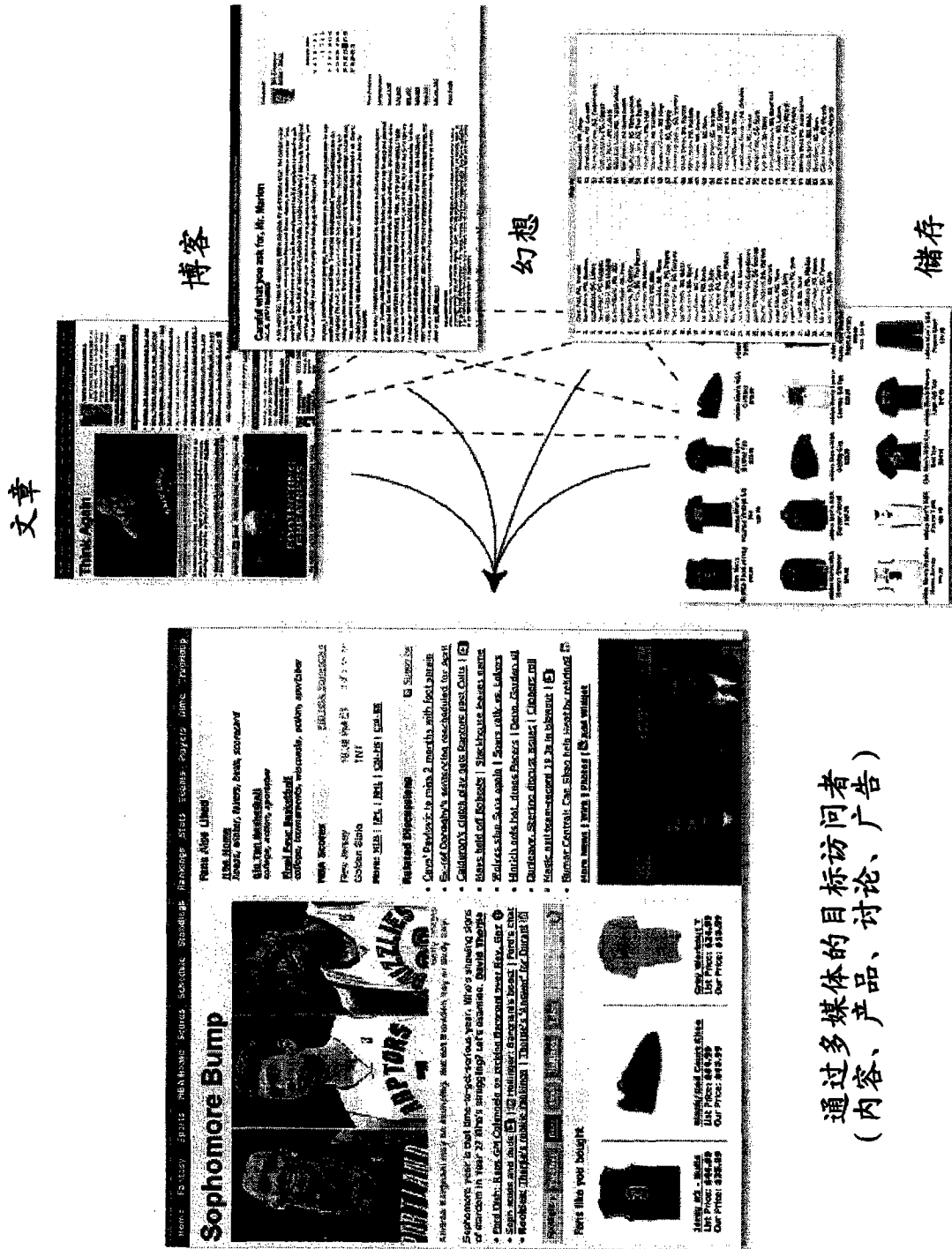


图 23

迷人的  
自维持  
兴趣驱动

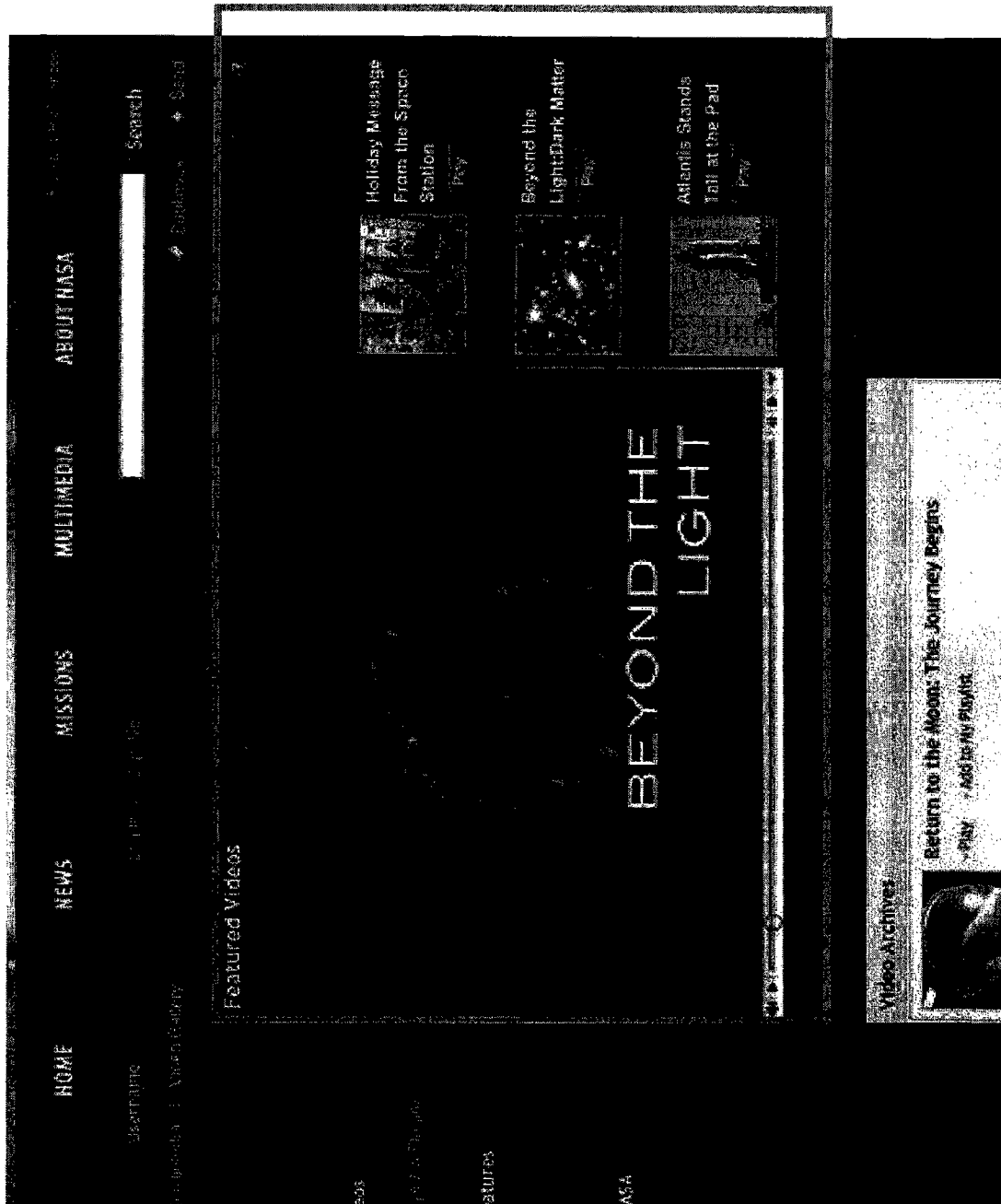
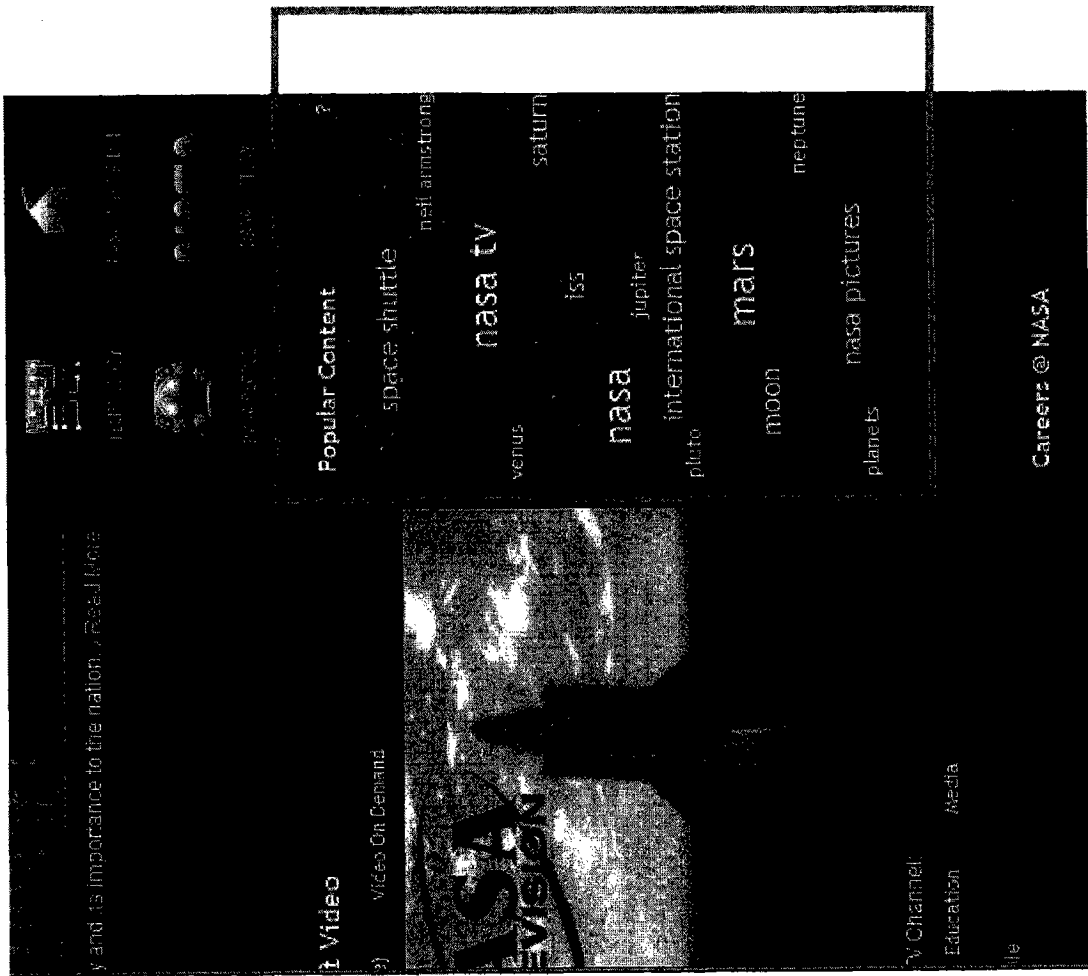


图 24



现在最迷人的“主题”

驱动内容发现

都隐性，不要求用户加标记

图 25

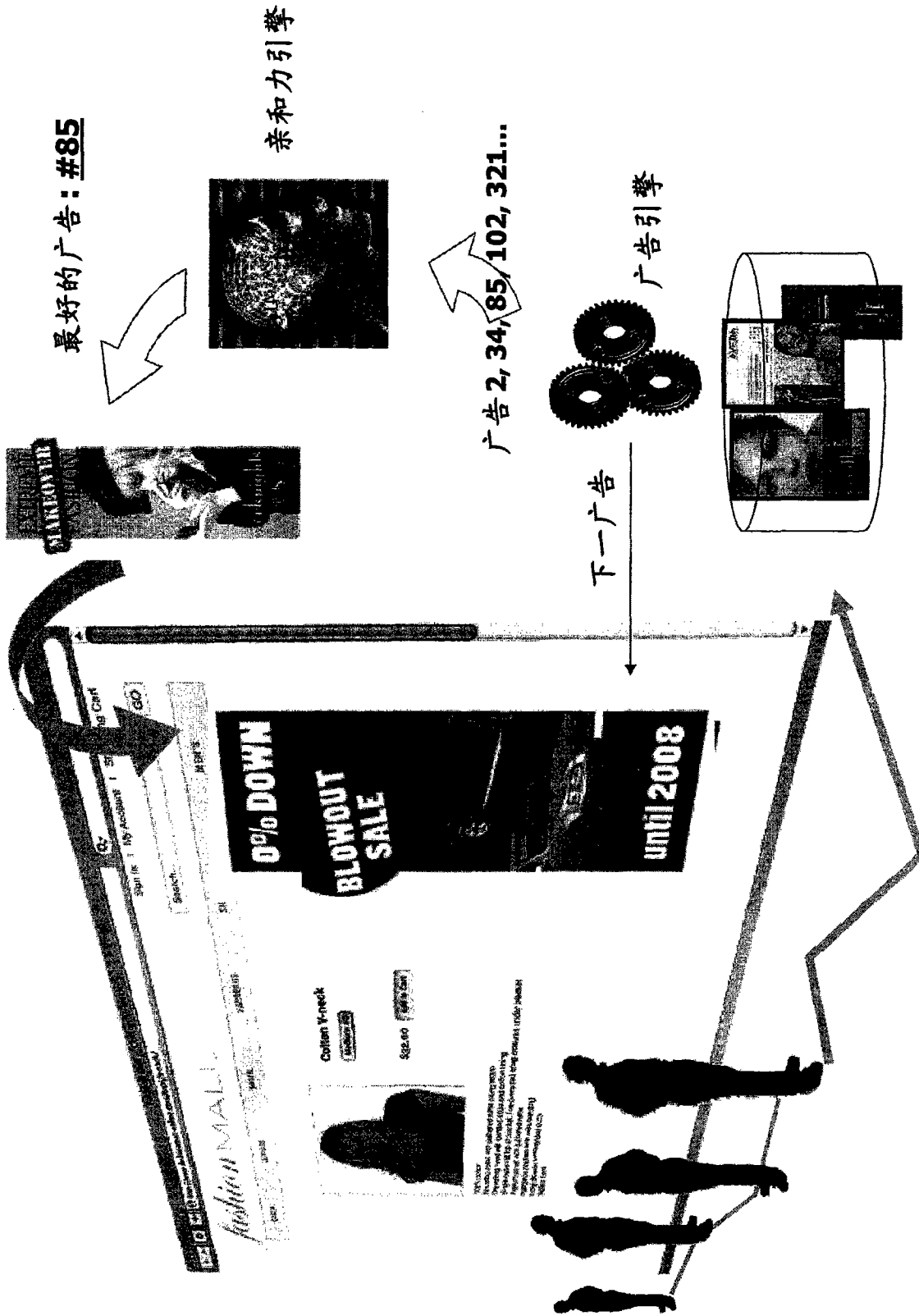


图 26

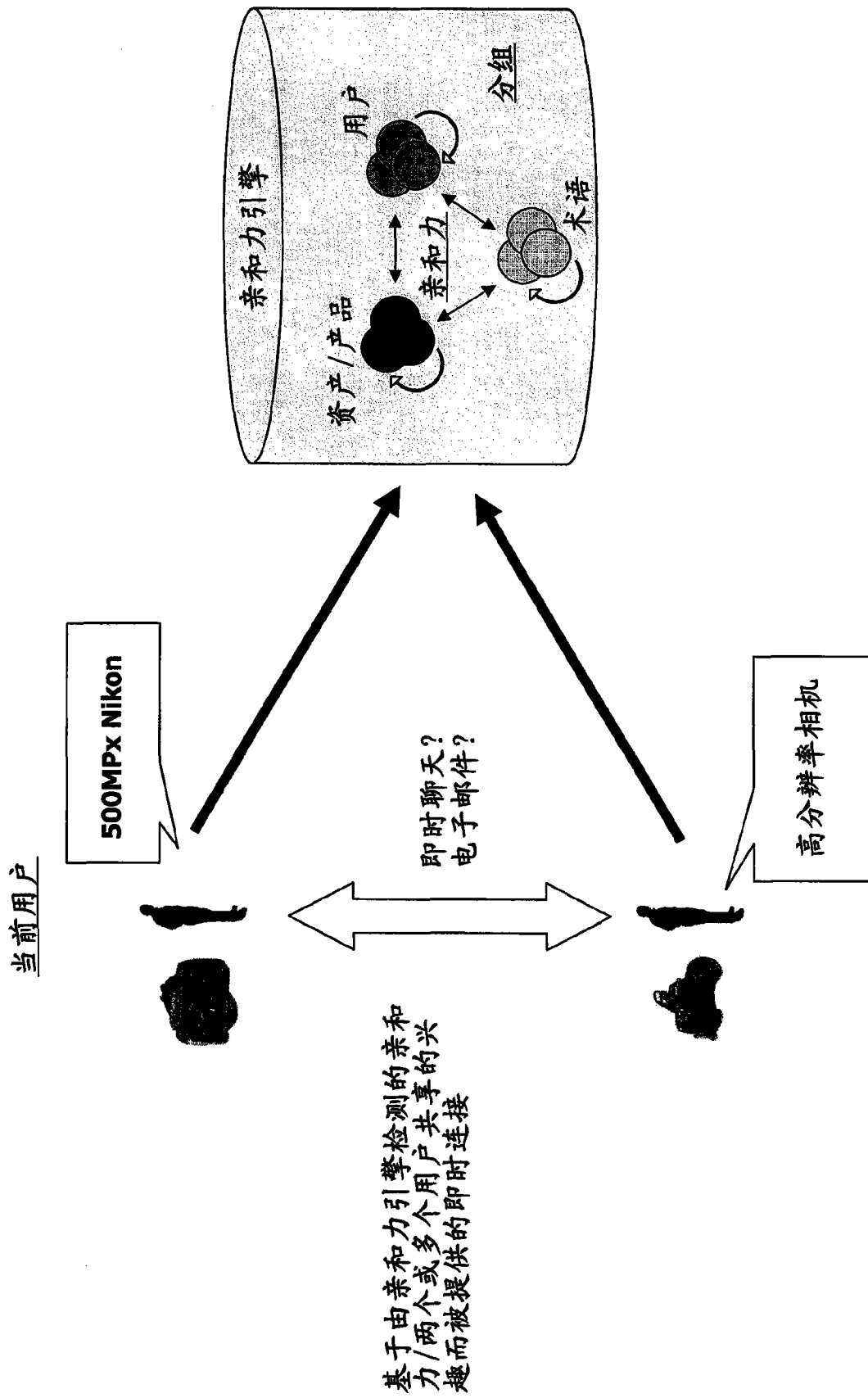


图 27

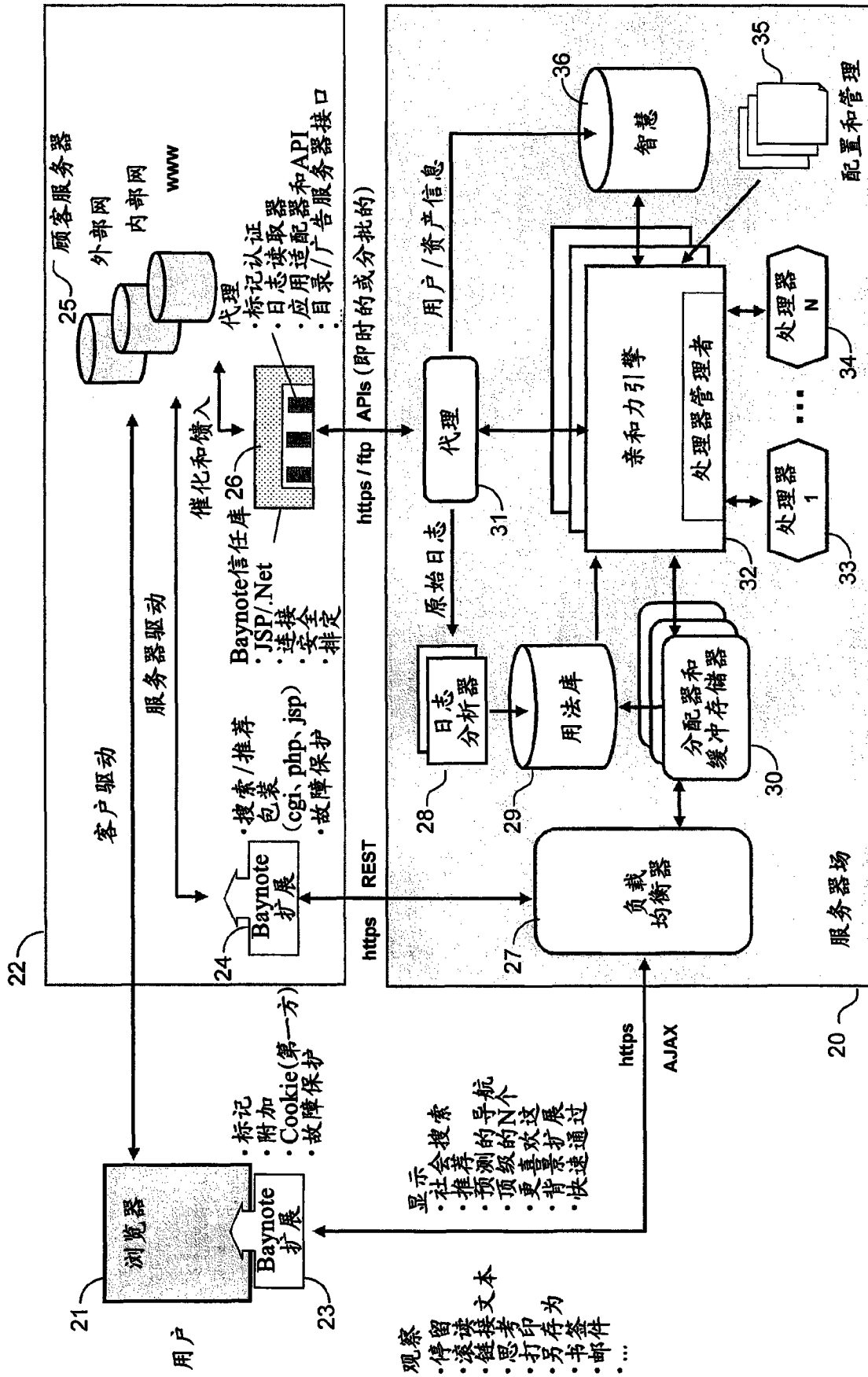


图 28

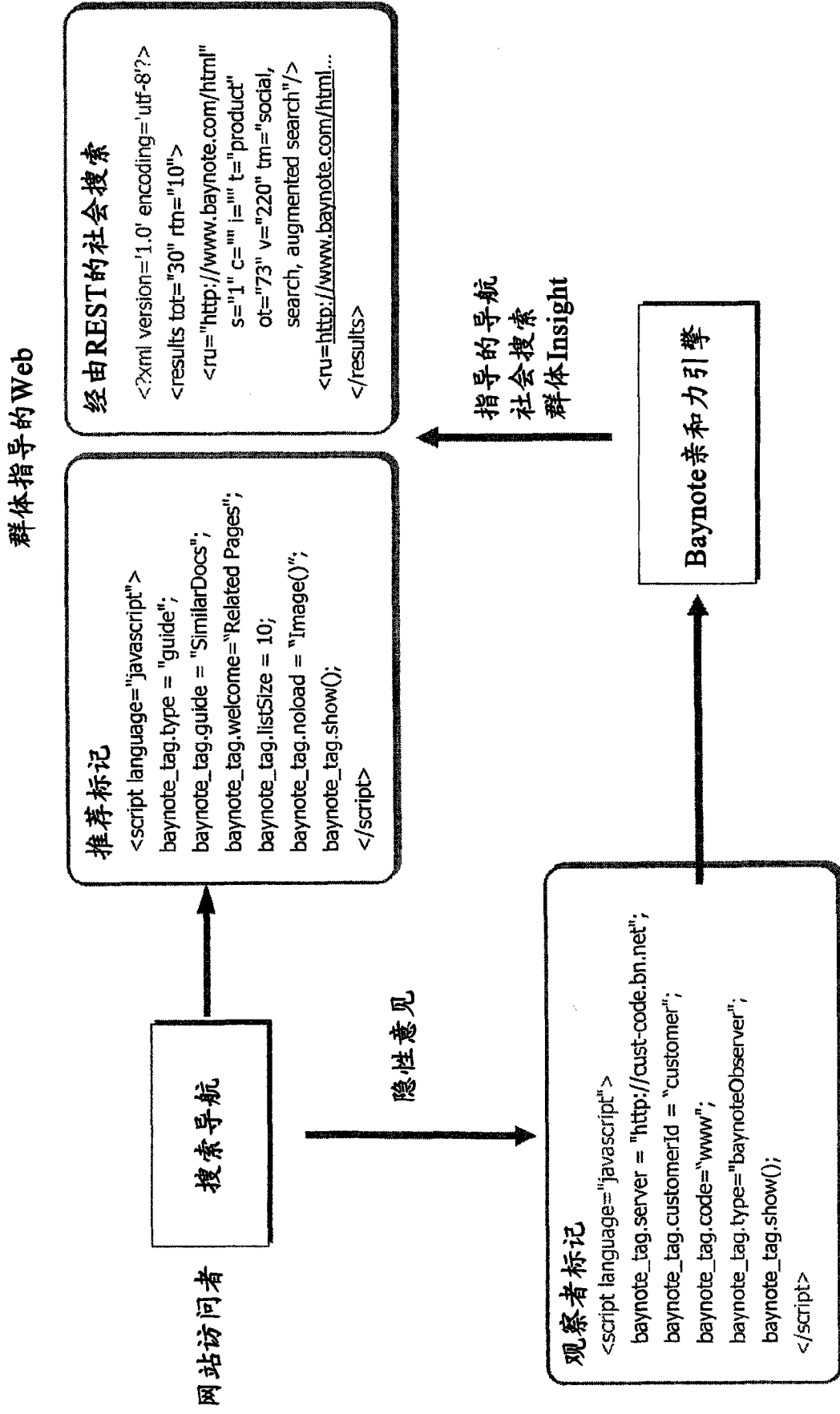


图 29



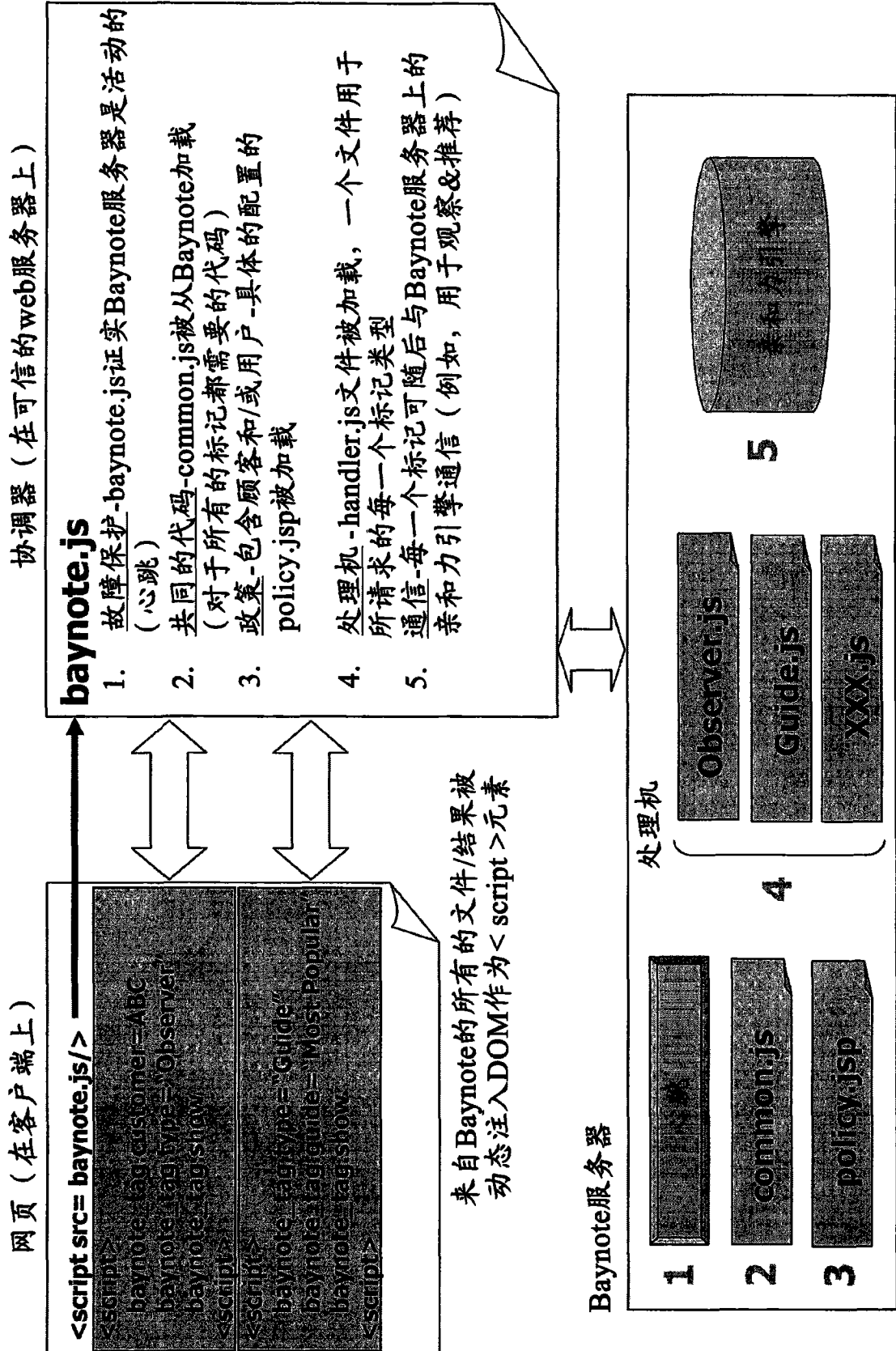


图 30

1. 一种用于内容推荐的由计算机实施的方法,包括以下步骤:

所述计算机接收用户动作所导致的信息,并识别当前用户的背景;

基于所述识别,所述计算机找到适当的对等组,所述对等组代表在所述识别的背景中与所述当前用户最像的其他用户;

所述计算机找到所述对等组识别为与所述当前背景最相关的内容;以及

所述计算机产生输出,所述输出包括用于显示给所述当前用户的推荐的内容;

所述计算机通过以下操作产生全面的行为指纹:

考虑包括多种多样的用户行为的输入;

将所有的这些用户行为与对应于所述当前用户在其他页面上的行为的储存信息进行互相关,并与群体中其余部分和识别出的对等体在当前页面上的行为进行互相关;

根据此分析,所述计算机产生所述当前用户在一段特定的内容中已经找到价值的可能性;以及

所述计算机储存所述可能性供在学习系统中使用。

2. 一种由计算机实施的以背景为中心的内容推荐方法,其包括以下步骤:

所述计算机基于代表网站上的用户动作的输入,为所述网站上的用户建立当前背景;

所述计算机用嵌入所述网站中的观察者标记捕获所述用户的当前背景信息;

所述计算机基于由所述观察者标记所收集的所有的用户行为确定所述用户的兴趣;

所述计算机使用全面行为指纹技术来分析所述行为;

所述计算机将所有的结果信息作为所述用户的当前背景矢量储存在存储器中,所述背景矢量包括在每一个条目上都带有权重的、术语和文档的混合矢量,所述权重反映该术语或文档反映所述用户的当前背景的强烈程度;

当用户输入搜索术语和/或点击导航链接时,所述计算机增加对应于被输入或被点击的术语和短语的所述背景矢量的条目,以捕获所表达的兴趣;

当用户动作进一步移动成为过去时,所述计算机减少或衰减相应的条目;

所述计算机增加用户点击的或基于用户的隐性动作所确定的用户表示感兴趣的相应的矢量条目文档;以及

根据前述操作,所述计算机产生所述用户的当前背景的代表来作为背景矢量,用于对所述用户进行内容推荐。

3. 如权利要求2所述的方法,还包括以下步骤:

基于从聚合智慧所学习到的亲和力和联系,所述计算机将所述背景矢量扩展和改进为意图矢量,所述聚合智慧从对群体作为一个整体的长期的观察中收集;

其中所述意图矢量通过查看在所述背景矢量中的术语和文档与其他术语的亲和力来创建。

4. 如权利要求3所述的方法,还包括以下步骤:

所述计算机利用亲和力引擎确定亲和力,所述亲和力允许将所述背景矢量扩展为意图矢量;

其中所述亲和力引擎通过观看用户群体在网站上的所有的隐性行为并通过应用行为指纹和使用等级算法来确定兴趣和联系,从而学习文档和术语、文档和文档、术语和术语、以及用户到其他用户、文档、和术语之间的连接。

5. 如权利要求 4 所述的方法,还包括以下步骤:

所述计算机识别共享对当前意图的亲活力的用户的对等组、以及在该意图的背景中展示出与所述当前用户最像的行为的那些用户,所述对等组由用户矢量代表,其中在所述用户矢量中的每一个用户条目可具有权重,该权重指示他在当前背景中对于所述当前用户是对等体的强烈程度。

6. 如权利要求 5 所述的方法,还包括以下步骤:

所述计算机识别哪些文档在当前意图中对所识别的对等体具有最高的亲和力;

其中所述亲和力引擎查看与那些对等体相关的、根据他们的对等体权重被加权的活动性矢量、术语 - 文档矩阵以及下一步矩阵,并聚合地使用这些来计算对所述当前意图具有最高的亲和力的那些文档;以及

其中所述文档是未过滤的推荐。

7. 如权利要求 6 所述的方法,还包括以下步骤:

所述计算机向所述亲和力引擎请求关于所述推荐中的每一个推荐的群体信息;

其中所述信息与其他资产信息结合并被显示给用户以帮助其了解推荐所隐含的群体智慧。

8. 如权利要求 7 所述的方法,其中所显示的信息包括以下中的任意项:在文档中找到价值的用户总数、将文档与当前背景和 / 或意图关联起来的对等体的数量、以及对等群体已经将之与一文档相关联的术语和 / 或短语。

9. 如权利要求 8 所述的方法,其中,所述术语和 / 或短语包括代表所述群体已经关联到一文档的术语的虚拟公众分类;

其中所述虚拟公众分类由所述亲和力引擎基于所述用户群体的隐性动作自动创建。

10. 一种用于内容推荐的设备,包括:

用于接收输入信息以在识别当前用户的背景中使用的装置;

用于基于所述识别而找到适当的对等组的装置,所述对等组代表在所述识别的背景中与所述当前用户最像的其他用户;

用于找到所述对等组识别为与所述当前背景最相关的内容的装置;

用于产生包括推荐给所述当前用户的内容的输出的装置;

用于通过以下操作产生全面行为指纹的装置:

考虑包括多种多样的用户行为的输入;

将所有的这些用户行为与所述当前用户在其他页面上的行为、以及与群体中其余部分和所识别的对等体在当前页面上的行为进行互相关;

根据此分析,产生所述当前用户在一段特定的内容中已经找到价值的可能性;以及  
储存所述可能性供学习系统使用。

11. 一种以背景为中心的内容推荐设备,包括:

用于接收代表用户动作的输入并用于根据所述输入在网站上建立当前用户背景的设备;

被嵌入所述网站用于捕获所述用户的当前背景信息的观察者标记;

用于基于由所述观察者标记所收集的各种的行为确定所述用户的兴趣的装置;

用于使用全面行为指纹技术来分析所述行为的装置;

用于将所有的结果信息作为所述用户的当前背景矢量储存在存储器中的装置,所述背景矢量包括在每一个条目上都带有权重的、术语和文档的混合矢量,所述权重反映该术语或文档反映所述用户的当前背景的强烈程度;

用于当用户输入搜索术语和 / 或点击导航链接时增加对应于被输入或被点击的术语和短语的所述背景矢量的条目以捕获所表达的兴趣的装置;

用于当用户动作进一步移动成为过去时减少或衰减相应的条目的装置;

用于增加用户点击的或基于用户的隐性动作所确定的用户表示感兴趣的相应的矢量条目文档的装置;以及

用于根据前述操作产生所述用户的当前背景的代表作为背景矢量用于对所述用户进行推荐的装置。

12. 如权利要求 11 所述的设备,还包括:

用于基于从聚合智慧所学习到的亲和力和联系将所述背景矢量扩展和改进为意图矢量的装置,所述聚合智慧从对群体作为一个整体的长期的观察中收集;

其中所述意图矢量通过查看在所述背景矢量中的术语和文档与其他术语的亲和力来创建。

13. 如权利要求 12 所述的设备,还包括:

用于利用亲和力引擎确定亲和力的装置,所述亲和力允许将所述背景矢量扩展为意图矢量;

其中所述亲和力引擎通过观看用户群体在网站上的所有的隐性行为并通过应用行为指纹和使用等级算法来确定兴趣和联系,从而学习文档和术语、文档和文档、术语和术语、以及用户到其他用户、文档、和术语之间的连接。

14. 如权利要求 13 所述的设备,还包括:

用于识别共享对当前意图的亲和力的用户的对等组以及在该意图的背景中展示出与所述当前用户最像的行为的那些用户的装置,所述对等组由用户矢量代表,其中在所述用户矢量中的每一个用户条目可具有权重,该权重指示他在当前背景中对于所述当前用户是对等体的强烈程度。

15. 如权利要求 14 所述的设备,还包括:

用于查看哪些文档在当前意图中对所识别的对等体具有最高的亲和力的装置;

其中所述亲和力引擎查看与那些对等体相关的、根据他们的对等体权重被加权的活动性矢量、术语 - 文档矩阵以及下一步矩阵,并聚合地使用这些来计算对所述当前意图具有最高的亲和力的那些文档;以及

其中所述文档是未过滤的推荐。

16. 如权利要求 15 所述的设备,还包括:

用于向所述亲和力引擎请求关于所述推荐中的每一个推荐的群体信息的装置;

其中所述信息与其他资产信息结合并被显示给所述用户以帮助其了解推荐所隐含的群体智慧。

17. 如权利要求 16 所述的设备,其中所显示的信息包括以下中的任意项:在文档中找到价值的用户总数、将文档与当前背景和 / 或意图关联起来的对等体的数量、以及对等群体已经将之与一文档相关联的术语和 / 或短语。

18. 如权利要求 17 所述的设备,其中,所述术语和 / 或短语包括代表所述群体已经关联到一文档的术语的虚拟公众分类;

其中所述虚拟公众分类由所述亲和力引擎基于所述用户群体的隐性动作自动创建。