

(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
2012 年 11 月 1 日 (01.11.2012)



(10) 国际公布号
WO 2012/146136 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2012/074093
- (22) 国际申请日: 2012 年 4 月 16 日 (16.04.2012)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201110107123.1 2011 年 4 月 26 日 (26.04.2011) CN
- (71) 申请人 (对除美国外的所有指定国): 北京百度网讯科技有限公司 (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): 闫鹏 (YAN, Peng) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 李彦宏 (LI, Yanhong) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 蔡虎 (CAI, Hu) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大

厦 2 层, Beijing 100085 (CN)。 沈毅 (SHEN, Yi) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 李磊 (LI, Lei) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。

- (74) 代理人: 深圳市威世博知识产权代理事务所 (普通合伙) (CHINA WISPRO INTELLECTUAL PROPERTY LLP.); 中国广东省深圳市南山区高新区粤兴三道 8 号中国地质大学产学研基地中地大楼 A806, Guangdong 518057 (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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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,

[见续页]

(54) Title: METHOD AND SYSTEM FOR INFORMATION SEARCHING

(54) 发明名称: 信息搜索方法及系统

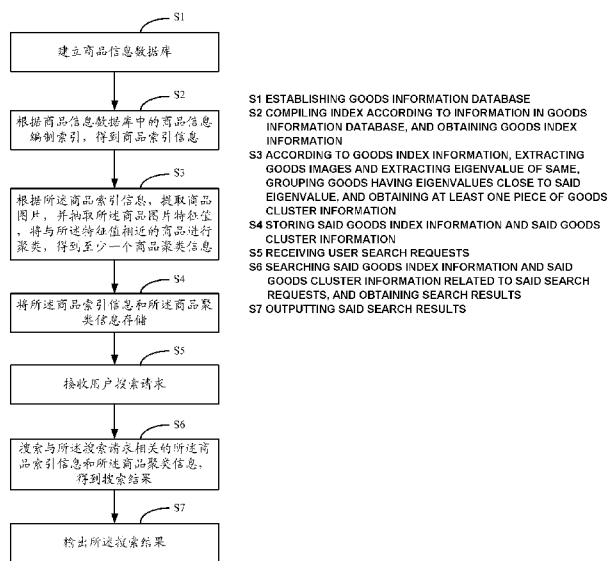


图 5 / Fig. 5

(57) Abstract: A method and system for information searching, said method comprising: compiling in advance an index according to goods information in a goods information database and obtaining goods index information; according to said goods index information, extracting goods images and extracting an eigenvalue for said images, grouping in clusters goods having eigenvalues close to said eigenvalue, and obtaining at least one item of goods cluster information; after receiving user search requests, searching said goods index information and said goods cluster information related to said search requests, and obtaining search results. The present invention clusters the same or similar goods, therefore reducing further user filtering operations, improving search efficiency, and saving network traffic, while also reducing the number of client requests for server access, reducing server processing stress, reducing network bandwidth occupation, improving network transmission speed, and avoiding network congestion.

(57) 摘要: 一种信息搜索方法和系统, 所述方法包括: 预先根据商品信息数据库中的商品信息编制索引, 得到商品索引信息; 根据所述商品索引信息, 提取商品图片, 并抽取所述商品图片特征值, 将与所述特征值相近的商品进行聚类, 得到至少一个商品聚类信息; 接收到用户搜索请求后, 搜索与所述搜索请求相关的所述商品索引信息和所述商品聚类

信息, 得到搜索结果。以上方案通过将相同或相似的商品进行聚类, 减少用户的进一步筛选操作, 提高搜索效率, 节约网络流量; 同时, 减少了客户端向服务器发起访问请求次数, 从而减轻了服务器的处理压力, 以及减少了占用的网络带宽, 提升了网络传输速度, 避免造成网络堵塞。



RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,

CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

说明书

发明名称：信息搜索方法及系统

- [1] 本申请要求了申请日为2011年04月26日，申请号为201110107123.1发明名称为“信息搜索方法及系统”的中国专利申请的优先权。
- [2] **【技术领域】**
- [3] 本发明涉及一种信息搜索方法及系统，尤其是指一种在搜索过程中可按照相似度进行聚类的搜索方法及系统。
- [4] **【背景技术】**
- [5] 随着互联网的普及以及在线支付技术的发展，通过互联网进行在线购物已经成为一种重要的购物方式，在线购物的优点主要包括：便捷，只要有互联网络的地方均可随时随地地进行购物；选择面广，可以同时浏览、比较众多不同商家出售的同类或者不同类的商品；价格便宜，相对于实体店面出售的商品，价格上存在较大的折扣。基于在线购物的这些优点，提供在线购物的在线购物平台已经得到了充分的发展。
- [6] 这些在线购物网站上都会提供一个商品搜索界面的输入框，用户在该搜索界面中输入商品查询条件时，会很快得到对应的商品搜索结果，这些商品搜索结果可以按照不同的属性进行排序，以便用户可以快速找到其最期望搜索到的商品，所述的排序属性包括：按照商品的销量排序、按照商品卖家的地域排序、按照商品卖家的信誉排序、按照商品的价格排序，等等。
- [7] 这些商品排序的方法有助于用户较快地查询到期望购买的商品。但是，当商品搜索结果很多时，以上的商品排序方法仍然不能完全满足用户的需要。常见的问题是：在现有商品搜索中，根据用户提交的商品查询关键字搜索出商品后，查询结果被直接罗列给用户使用，所述的商品搜索结果中，有很多是相同的商品，只是这些提供这些商品的卖家不同，即相同或类似的商品不进行聚类提供给用户使用，需要用户在定位自己真正期望的商品之前，进行多次的筛选或者搜索。以图1中的一种现有的商品搜索结果为例，用户搜索的关键词是“诺基亚手机”(对应于图1中的区域“A”)，由于诺基亚手机的型号众多、应用广泛、且卖家

众多，返回的搜集结果(对应于图1中的区域“B”，图1中仅仅显示了几个搜索结果)多达上万项，而在这上万项搜索结果中，有很多搜索结果对应的都是型号相同的手机，只是出售这些型号相同的手机商品的卖家不同而已，类似于图1中这样直接罗列搜索结果的缺陷在于：若用户期望详细比较察看其中某款型号的手机，他可能需要再次限定新的更具体的搜索关键词，或者他可以选择一页一页地翻看已有的大量检索结果，无论采用那种方式，用户都需要进行多次的筛选或者搜索，这样的商品搜索结果的展现方式不能完全符合用户的期望值，搜索效率较低，耗费无谓的网络流量，另外，无论用户是限定新的更具体的搜索关键词，还是用户不断地翻页、点击查看每条搜索结果，均使得客户端向服务器不断的发起访问请求，加大了服务器的处理压力；与此同时，服务器也需要不断的向客户端返回搜索结果，这些搜索结果在网络传输的过程中将占用大量的网络带宽，减慢了网络传输速度，甚至造成网络堵塞。

[8] 因此，有必要在现有技术的基础上做出改进，向在线购买商品的用户提供一种更准确的商品搜索结果的商品搜索方法及系统。

[9] **【发明内容】**

[10] 本发明的目的在于为解决上述现有技术缺陷，提供一种经过改进的在商品搜索中按照商品的相似度进行聚类的商品搜索方法，所述的商品搜索方法可以在商品搜索结果中将相同或者相近似的商品聚类组合在一起展现。

[11] 本发明的另一目的在于为解决上述现有技术缺陷，提供一种经过改进的在商品搜索中按照商品的相似度进行聚类的商品搜索系统，所述的商品搜索系统可应用所述经过改进的在商品搜索中按照商品的相似度进行聚类的商品搜索方法。

[12] 相应地，本发明一种实施方式的商品搜索方法包括如下步骤：

[13] 预先根据商品信息数据库中的商品信息编制索引，得到商品索引信息；根据所述商品索引信息，提取商品图片，并抽取所述商品图片特征值，将与所述特征值相近的商品进行聚类，得到至少一个商品聚类信息；

[14] 接收到用户搜索请求后，搜索与所述搜索请求相关的所述商品索引信息和所述商品聚类信息，得到搜索结果。

[15] 相应地，本发明的一种实施方式的商品搜索系统包括：

- [16] 聚类模块，用于预先根据商品信息数据库中的商品信息编制索引，得到商品索引信息；根据所述商品索引信息，提取商品图片，并抽取所述商品图片特征值，将与所述特征值相近的商品进行聚类，得到至少一个商品聚类信息；
- [17] UI模块，用于接收用户搜索请求；
- [18] 搜索模块，用于搜索与所述搜索请求相关的所述商品索引信息和所述商品聚类信息，得到搜索结果。
- [19] 本发明的有益效果是：改进了商品搜索的方法及系统，通过增加商品的聚类模块，优化搜索的排序结果，将相同或相似的商品进行聚类，增加商品列表的规整程度，减少用户的进一步筛选操作，提高搜索效率，节约网络流量；同时，减少了客户端向服务器发起访问请求次数，从而减轻了服务器的处理压力，以及减少了占用的网络带宽，提升了网络传输速度，避免造成网络堵塞。

[20] **【附图说明】**

- [21] 图1 是一种现有的在线购物网站的商品查询结果示意图。
- [22] 图2是本发明的商品搜索系统与客户端实现互动的工作原理图。
- [23] 图3是本发明的一种实施方式商品搜索系统的结构示意图。
- [24] 图4是图3中的聚类模块的结构示意图。
- [25] 图5是本发明的一种实施方式的商品搜索方法的工作流程示意图。
- [26] 图6是应用本发明的商品搜索方法所产生的商品查询结果按照聚类组合的网页示意图。
- [27] 图7是图6中的一个商品聚类展开后的商品查询结果网页示意图。

[28] **【具体实施方式】**

- [29] 为了对发明的技术特征、发明目的和技术效果有更加清楚的理解，现对照附图说明本发明的具体实施方式，在各图中相同的标号表示步骤相同的部分。在本文中，“示意性”表示“充当实例、例子或说明”，不应将在本文中被描述为“示意性”的任何图示、实施方式解释为一种更优选的或更具优点的技术方案。
- [30] 如图2所示的本发明的商品搜索中按照商品相似度聚类的商品搜索系统10与客户端20实现互动的工作原理图。本实施方式中，该客户端20包括一浏览器，客户可通过该浏览器打开搜索引擎，并在搜索引擎中输入查询条件，一般的，该

输入的查询条件为文本信息，当然，该查询条件还可以为图片信息、视频信息等等，当然，在本发明最佳实施方式中，所述搜索引擎即是本发明所述的商品搜索系统。所述商品搜索系统10通过网络接收客户输入至所述浏览器中的查询条件，并对该查询条件进行搜索后，将搜索结果返回至该浏览器。其中，该商品搜索系统10可以包括一台或多台服务器，该客户端20可以包括一个或多个用户终端设备，如个人计算机、笔记本电脑、无线电话、个人数字处理（PDA）、或其它计算机系统和通信系统。

[31] 这些服务器和终端设备在架构上都包含一些基本组件，如总线、处理系统、存储系统、一个或多个输入/输出系统、和通信接口等。总线可以包括一个或多个导线，用来实现服务器或终端设备各组件之间的通信。处理系统包括各类型的用来执行指令、处理进程或线程的处理器或微处理器。存储系统可以包括存储动态信息的随机访问存储器（RAM）等动态存储器，和存储静态信息的只读存储器（ROM）等静态存储器，以及包括磁或光学记录介质与相应驱动的大容量存储器。输入系统供用户输入信息到服务器或终端设备，如键盘、鼠标、手写笔、声音识别系统、或生物测定系统等。输出系统包括用来输出信息的显示器、打印机、扬声器等。通信接口用来使服务器或终端设备与其它系统或系统进行通信。通信接口之间可通过有线连接、无线连接、或光连接连接到网络中，使商品搜索系统10、客户端20间能够通过网络实现相互间的通信。网络可以包括局域网（LAN）、广域网（WAN）、电话网络如公共交换电话网（PSTN）、企业内部的互联网、因特网、或上述这些网络的结合等。

[32] 服务器和终端设备上均包含有用来管理系统资源、控制其它程序运行的操作系统软件，以及用来实现特定功能模块的应用软件。如图3所示，本发明的一种实施方式的商品搜索系统包括web服务模块100、与web服务模块100交互通信的UI模块102、与所述UI模块交互通信的搜索模块104、与所述搜索模块104通信的聚类模块106、与所述聚类模块106通信的数据存储模块108，以及与所述数据存储模块108通信的更新模块109。值得一提的是，这些模块即可存储并运行于同一服务器中，也可存储并运行在多台服务器中。

[33] 所述web服务模块100用于通过网络协议接收从客户端20传来的查询条件，并将

该查询条件转到UI模块102，另外，该web服务模块100还用于接收所述UI模块102返回的结果页面，并将所述结果页面返回至客户端20。

[34] 所述UI模块102用于接收所述web服务模块100传送的查询条件，并将该查询条件发送至所述搜索模块104；另外，所述UI模块102还用于接收所述搜索模块104返回的搜索结果，并将所述搜索结果拼装为结果页面后，返回所述结果页面至所述web服务模块100。

[35] 所述搜索模块104用于搜索与所述搜索请求相关的商品索引信息、商品聚类信息，得到搜索结果。具体地，所述搜索模块104接收所述UI模块102输入的查询条件，并根据所述查询条件在所述聚类模块106中查询，得到包括至少一个商品索引信息，以及商品聚类信息的搜索结果。优选地，在本发明最佳实施方式中，所述搜索模块104还用于根据聚类模块106中的聚类情况，搜索得到聚类商品平均价格信息，并将所述结果发送至所述UI模块102。

[36] 所述聚类模块106用于建立并存储商品索引信息、商品聚类信息。在本发明最佳实施方式中，所述聚类模块106还用于建立并存储商品平均价格信息。如何得到所述商品索引信息、商品聚类信息、商品平均价格信息可参下述结合图5的具体描述。

[37] 所述数据存储模块108用于建立商品信息数据库。具体地，所述数据存储模块108存储大量卖家提交的商品信息，本发明所用到的商品信息中主要是商品名称与商品的图片。当然，优选地，所述数据存储模块108中的商品信息还包括商品价格、类别、名称、摘要等等。

[38] 所述更新模块109用于定期根据商品信息数据库的更新信息，更新所述商品索引信息、所述商品聚类信息。优选地，所述更新模块109还用于定期更新所述商品平均价格信息。

[39] 如图4所示，在本发明一个实施方式中，所述聚类模块106包括索引单元1060、聚类单元1062、价格计算单元1064、以及存储单元1066。

[40] 索引单元1060用于根据所述商品信息数据库中的商品名称编制索引，得到商品索引信息。优选地，在本发明最佳实施方式中，该索引不仅可将存储于数据存储模块108中所有的商品信息建立索引，还可按照商品属性将索引分成多个对应

的类别，例如：若按照商品属性为品牌建立索引，则检索所有商品名称中带有品牌的，如将商品名称中包含“诺基亚”归入一类并建立对应的索引，将商品名称中包含“摩托罗拉”的归入另一类并对应的建立索引，以此类推，即可建立按照商品属性为品牌的完整索引表，当然，该商品属性还可以为其他，如品种、销售区域、销售价格、卖家信用等等，按照上述商品属性为品牌相类似的方式，可以为储存于数据存储模块108中的所有商品建立一个完整的索引表。

[41] 聚类单元1062用于根据所述商品索引信息，提取商品图片，并抽取所述商品图片特征值，将与所述特征值相近的商品进行聚类，得到至少一个商品聚类信息。所述的商品的图片指的是存储在数据存储模块108中的所有商品的商品信息中包括的表示该商品特征的至少一张图片信息。所述的商品的特征值通过SIFT算法进行计算获得，通过提取图片的局部特征，在尺度空间寻找极值点，提取位置，尺度，旋转不变量，并去掉尺度变化，旋转等几何变形因素的影响，其主要的步骤包括：1. 检测尺度空间极值点；2. 精确定位极值点；3. 为每个关键点指定方向参数；4. 关键点描述子的生成。关于SIFT算法的详细介绍可以参考公开资料，这里不再详细描述。在通过SIFT算法进行特征值计算后，对进行量化，将量化之后的特征值量值相同或相似的图片进行聚类。

[42] 价格计算单元1064用于计算出至少一个商品聚类信息中的商品平均价格信息。一般而言，即使对于一个商品聚类中的同样商品，由于卖家的不同，商品的销售价格会存在一定的区别，通过统计某个商品聚类信息中的所有商品的价格并获得一个整体的平均价格，该平均价格可以作为搜索结果的一部分提供给用户参考，让用户可以更清晰地了解该商品的销售行情，有利于提高客户的使用体验度。

[43] 存储单元1066用于存储所述商品索引信息和商品聚类信息。优选地，在本发明最佳实施方式中，所述存储单元1066还用于存储商品平均价格信息，以便搜索模块104可以迅速的搜索所述存储单元1066中的商品索引信息、商品聚类信息、商品平均价格信息，并将搜索到的结果通过UI模块102拼装为结果页面，传输给用户。

[44] 以下结合图5描述本发明的一种实施方式的商品搜索中按照商品的相似度聚类

的方法。

[45] 图5是本发明的一种实施方式的商品搜索方法的工作流程示意图。所述的方法包括如下的步骤：

[46] S1、建立商品信息数据库；所述商品信息数据库储存于如图3所述的数据存储模块108中。所述商品信息主要包括商品名称与商品的图片，当然，优选地，所述商品信息还可包括商品价格、类别、名称、摘要等。所述建立商品信息数据库的步骤，本领域普通技术人员已可通过现有技术熟练掌握，在此不再赘述。

[47] S2、根据商品信息数据库中的商品名称编制索引，得到商品索引信息；优选地，在本发明最佳实施方式中，该索引不仅可将存储于数据存储模块108中所有的商品信息建立索引，还可按照商品属性将索引分成多个对应的类别，例如：若按照商品属性为品牌建立索引，则检索所有商品名称中带有品牌的，如将商品名称中包含“诺基亚”归入一类并建立对应的索引，将商品名称中包含“摩托罗拉”的归入另一类并对应的建立索引，以此类推，即可建立按照商品属性为品牌的完整索引表，当然，该商品属性还可以为其他，如品种、销售区域、销售价格、卖家信用等等，按照上述商品属性为品牌相类似的方式，可以为储存于数据存储模块108中的所有商品建立一个完整的索引表。在本发明的一个实施方式中，所述的步骤S200通过索引单元1060完成。

[48] S3、根据所述商品索引信息，提取商品图片，并抽取所述商品图片特征值，将与所述特征值相近的商品进行聚类，得到至少一个商品聚类信息；在本发明的一个实施方式中，所述的步骤S3通过图5中的聚类单元1062完成，聚类单元1062用于比较处于某个商品索引分类下的商品的图片的特征值，将图片的特征值相近的商品放入一个聚类中。所述的商品的图片指的是存储在数据存储模块108中的所有商品的商品信息中包括的表示该商品特征的至少一张图片信息。所述的商品的特征值通过SIFT算法进行计算获得，通过提取图片的局部特征，在尺度空间寻找极值点，提取位置，尺度，旋转不变量，并去掉尺度变化，旋转等几何变形因素的影响，主要的步骤包括：1. 检测尺度空间极值点；2. 精确定位极值点；3. 为每个关键点指定方向参数；4. 关键点描述子的生成。关于SIFT算法的详细介绍可以参考公开资料，这里不再详细描述。在通过SIFT算法进行

特征值计算后，对进行量化，将量化之后的特征值量值相同或相似的图片进行聚类。

- [49] S4、将所述商品索引信息和所述商品聚类信息存储；这些商品索引信息、商品聚类信息均存储在所述的聚类模块106的存储单元1066中，以便搜索模块104可以迅速的搜索所述存储单元1066中的商品索引信息、商品聚类信息并将搜索到的结果通过UI模块102拼装为结果页面，传输给用户。优选地，所述聚类模块106定期读取数据存储模块108中商品信息数据库的变化情况并将其更新到索引聚类存储模块中，即是定期根据数据存储模块108中商品信息数据库的更新信息，更新所述商品索引信息、商品聚类信息。
- [50] S5、接收用户搜索请求；所述的搜索请求为用户在客户端浏览中输入的商品查询条件，该关键词通过Web服务模块100、UI模块102传输给所述搜索模块104。
- [51] S6、搜索与所述搜索请求相关的所述商品索引信息和所述商品聚类信息，得到搜索结果。优选地，在本发明的最佳实施方式中，在该步骤中，还有搜索与所述搜索请求相关的商品平均价格信息。另外，在本发明的一个实施方式中，所述步骤S2是通过图3中的搜索模块104完成。所述搜索单元1030用于接收所述UI模块102输入的查询条件，并将所述查询条件在所述聚类模块106中查询，得到聚类后的至少一类商品，以及与所述商品相应的商品平均价格。
- [52] S7、输出搜索结果。所述的搜索结果通过UI模块102拼装为结果页面后，通过Web服务模块100发送至用户客户端。
- [53] 在本发明最佳实施方式中，在上述S3步骤后，还包括步骤（图中未示出）：计算出至少一个所述商品聚类信息中的商品平均价格信息。在本发明的一个实施方式中，该步骤通过价格计算单元1064完成，价格计算单元1064用于计算出所述的一个聚类中的商品平均价格。一般而言，即使对于一个聚类中的用样商品，由于卖家的不同，商品的销售价格会存在一定的区别，通过统计某个聚类中的所有商品的价格并获得一个整体的平均价格，该平均价格可以作为搜索结果的一部分提供给用户参考，让用户可以更清晰地了解该商品的销售行情，有利于提高客户的使用体验度。
- [54] 相应地，在所述步骤S4中，还包括将所述商品平均价格信息进行存储。所述聚

类模块106定期读取数据存储模块108中商品信息数据库的变化情况并将其更新到索引聚类存储模块中，即是定期根据数据存储模块108中商品信息数据库的更新信息，更新所述商品索引信息、商品聚类信息、商品平均价格信息。

[55] 相应地，在所述步骤S6中，还包括搜索与所述搜索请求相关的商品平均价格信息。

[56] 参考图6和图7，以下举例说明本发明的商品搜索方法及系统的工作过程与搜索结果。

[57] 在图6中，通过在客户端浏览器打开的商品搜索引擎的查询窗口(图6中标示为“C”的区域)中输入查询条件“诺基亚 手机”，然后经过搜索，得到的聚类查询结果经过步骤通过UI模块102拼装成结果页面后，由Web服务模块发回客户端浏览器显示出来，如图6中标示为“D1”、“D2”、“D3”的区域分别显示了三种型号的诺基亚手机的聚类结果，所述的聚类结果中包括了商品的名称、简介、图片等，同时，还包括商品平均价格，以及商品的销售商家的数量。

[58] 图7是点击图6中的“D1”商品聚类展开后的商品查询结果示意图。可以看到，在将图6中的“D1”商品聚类展开之后，在该聚类中的所有商品的型号都是相同的(即型号为5800的诺基亚手机)，所有的商品仅按照卖家的不同(图7中标示为“E”的区域)逐一提供给用户。

[59] 通过以上描述，可以得出，本发明的商品搜索方法及系统，改进了商品搜索的方法及系统，通过增加商品的聚类模块，优化搜索的排序结果，将相同或相似的商品进行聚类，增加商品列表的规整程度，减少用户的进一步筛选操作，提高搜索效率，节约网络流量；同时，减少了客户端向服务器发起访问请求次数，从而减轻了服务器的处理压力，以及减少了占用的网络带宽，提升了网络传输速度，避免造成网络堵塞。

[60] 应当理解，虽然本说明书按照实施例加以描述，但并非每个实施例仅包含一个独立的技术方案，说明书的这种叙述方式仅仅是为清楚起见，本领域技术人员应当将说明书作为一个整体，各实施例中的技术方案也可以经适当组合，形成本领域技术人员可以理解的其他实施方式。

[61] 上文所列出的一系列详细说明仅仅是针对本发明的可行性实施例的具体说明

，它们并非用以限制本发明的保护范围，凡未脱离本发明技艺精神所作的等效实施例或变更均应包含在本发明的保护范围之内。

权利要求书

- [权利要求 1] 一种信息搜索方法，其特征在于，该方法包括如下步骤：
预先根据商品信息数据库中的商品信息编制索引，得到商品索引信息；根据所述商品索引信息，提取商品图片，并抽取所述商品图片特征值，将与所述特征值相近的商品进行聚类，得到至少一个商品聚类信息；
接收到用户搜索请求后，搜索与所述搜索请求相关的所述商品索引信息和所述商品聚类信息，得到搜索结果。
- [权利要求 2] 如权利要求1所述的方法，其特征在于，所述方法还包括：
定期根据商品信息数据库的更新信息，更新所述商品索引信息、商品聚类信息。
- [权利要求 3] 如权利要求1所述的方法，其特征在于，在得到至少一个商品聚类信息之后，还包括：
计算出至少一个所述商品聚类信息中的商品平均价格信息。
- [权利要求 4] 如权利要求3所述的方法，其特征在于，还包括：
在接收到用户搜索请求后，搜索与所述搜索请求相关的所述商品平均价格信息。
- [权利要求 5] 如权利要求4所述的方法，其特征在于，还包括：
定期根据商品信息数据库的更新信息，更新所述商品平均价格信息。
- [权利要求 6] 如权利要求1所述的方法，其特征在于，所述商品图片特征值计算是通过SIFT算法进行。
- [权利要求 7] 如权利要求1所述的方法，其特征在于，所述商品索引信息包括商品图片、商品价格、商品名称和/或商品型号。
- [权利要求 8] 一种信息搜索系统，其特征在于，所述系统包括：
聚类模块，用于预先根据商品信息数据库中的商品信息编制索引，得到商品索引信息；根据所述商品索引信息，提取商品图片，并抽取所述商品图片特征值，将与所述特征值相近的商品进行聚

类，得到至少一个商品聚类信息；
UI模块，用于接收用户搜索请求；
搜索模块，用于搜索与所述搜索请求相关的所述商品索引信息和所述商品聚类信息，得到搜索结果。

[权利要求 9] 如权利要求8所述的系统，其特征在于，所述系统还包括：
更新模块，用于定期根据商品信息数据库的更新信息，更新所述商品索引信息、所述商品聚类信息。

[权利要求 10] 如权利要求8所述的系统，其特征在于，所述聚类模块还用于：
计算出至少一个所述商品聚类信息中的商品平均价格信息。

[权利要求 11] 如权利要求10所述的系统，其特征在于，所述搜索模块还用于搜索与所述搜索请求相关的所述商品平均价格信息。

[权利要求 12] 如权利要求9所述的系统，其特征在于，所述更新模块还用于：
用于定期根据商品信息数据库的更新信息，更新所述商品索引信息、所述商品聚类信息，以及所述商品平均价格信息。

[权利要求 13] 如权利要求8所述的系统，其特征在于，所述商品图片特征值计算是通过SIFT算法进行。

[权利要求 14] 如权利要求8所述的系统，其特征在于，所述系统还包括：
Web服务模块，用于通过网络协议接收客户端发出的查询条件，并将所述查询条件转到所述UI模块，且所述web服务模块还用于接收所述UI模块返回的结果页面，并将所述结果页面返回至所述客户端。

[权利要求 15] 如权利要求8中任意一项所述的系统，其特征在于，所述商品索引信息包括商品图片和商品价格，以及商品名称和/或商品型号。

说明书附图



图 1

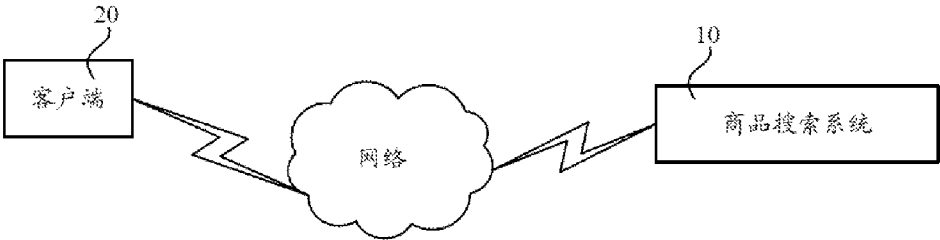


图 2

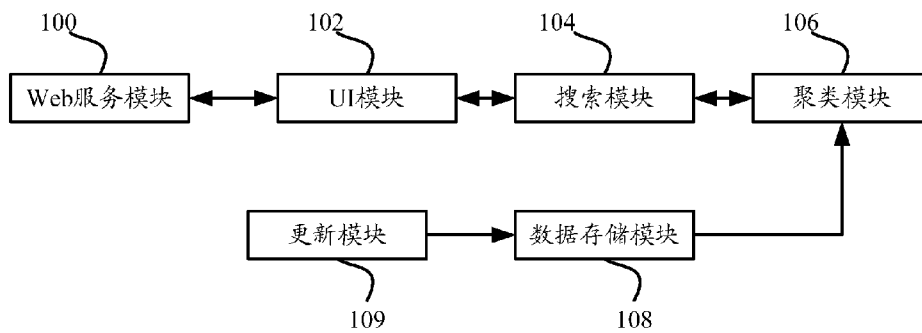


图 3

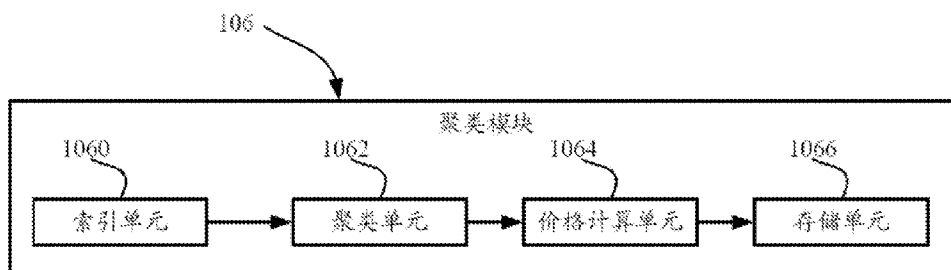


图 4

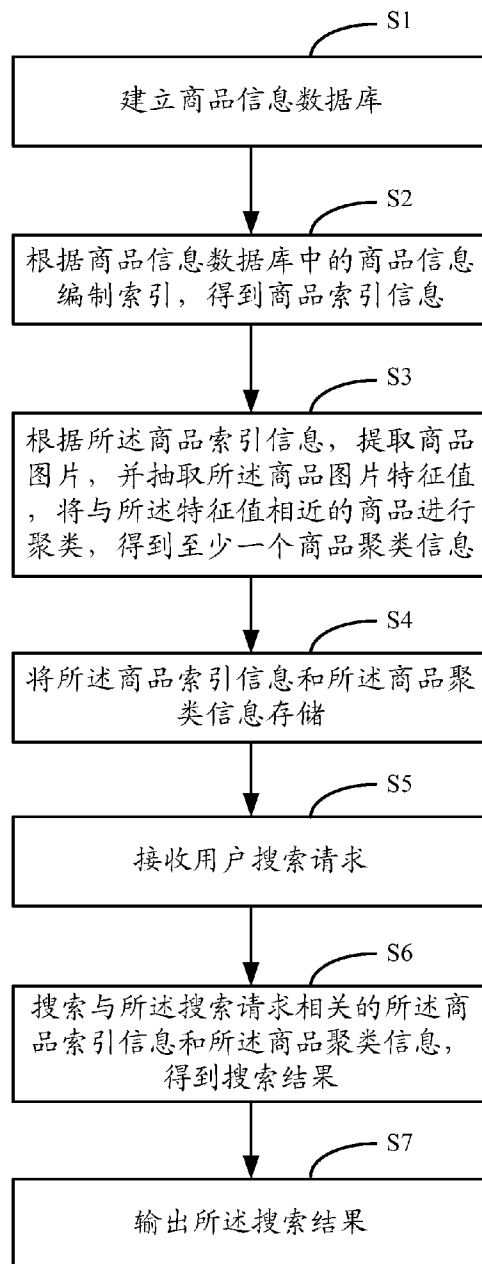


图 5



图 6



图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/074093

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)

CNPAT, WPI, EPODOC: information, search, database, index, eigenvalue, cluster, request, update, extract

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101853299 A (HANGZHOU TAOTAOSOU TECHNOLOGY CO, LTD.), 06. October 2010 (06.10.2010), description, paragraphs 32-48 and 53-67, and figure 1	1-15
A	CN 101211355 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES), 02 July 2008 (02.07.2008), the whole document	1-15
A	CN 101650717 A (ALIBABA GROUP HOLDING LTD.), 17 February 2010 (17.02.2010), the whole document	1-15
A	US 2008/0077569 A1 (YAHOO! INC.), 27 March 2008 (27.03.2008), the whole document	1-15

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“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)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search

04 July 2012 (04.07.2012)

Date of mailing of the international search report

02 August 2012 (02.08.2012)

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

CHEN, LongTelephone No.: (86-10) **62414059**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2012/074093

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101853299 A	06.10.2010	None	
CN 101211355 A	02.07.2008	None	
CN 101650717 A	17.02.2010	WO 2010019220 A1	18.02.2010
		US 2010042657 A1	18.02.2010
		EP 2316073 A1	04.05.2011
		JP 2011530767 A	22.12.2011
		US 2012124004 A1	17.05.2012
US 2008/0077569 A1	27.03.2008	KR 20080028574 A	01.04.2008

国际检索报告

国际申请号

PCT/CN2012/074093

A. 主题的分类

G06F 17/30(2006.01)i

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

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: G06F, H04L

包含在检索领域中的除最低限度文献以外的检索文献

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))

CNPAT, WPI, EPODOC: 信息, 搜索, 数据库, 索引, 特征值, 聚类, 请求, 更新, 提取, information, search, database, index, eigenvalue, cluster, request, update, extract

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN101853299A(杭州淘淘搜科技有限公司) 06. 10 月 2010(06.10.2010) 说明书第 32-48、53-67 段、图 1	1-15
A	CN101211355A(中国科学院计算技术研究所) 02. 7 月 2008(02.07.2008) 全文	1-15
A	CN101650717A(阿里巴巴集团控股有限公司)17. 2 月 2010(17.02.2010) 全文	1-15
A	US2008/0077569A1(YAHOO! INC.) 27. 3 月 2008(27.03.2008) 全文	1-15



其余文件在 C 栏的续页中列出。



见同族专利附件。

* 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

“&” 同族专利的文件

国际检索实际完成的日期

04.7 月 2012(04.07.2012)

国际检索报告邮寄日期

02.8 月 2012 (02.08.2012)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

陈龙

电话号码: (86-10) **62414059**

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2012/074093

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN101853299A	06.10.2010	无	
CN101211355A	02.07.2008	无	
CN101650717A	17.02.2010	WO2010019220A1	18.02.2010
		US2010042657A1	18.02.2010
		EP2316073A1	04.05.2011
		JP2011530767A	22.12.2011
		US2012124004A1	17.05.2012
US2008/0077569A1	27.03.2008	KR20080028574A	01.04.2008