



(12)发明专利申请

(10)申请公布号 CN 107679091 A

(43)申请公布日 2018.02.09

(21)申请号 201710799147.5

(22)申请日 2017.09.07

(71)申请人 北京海融兴通信息安全技术有限公司

地址 100091 北京市海淀区闵庄路3号102
幢

(72)发明人 白云飞

(74)专利代理机构 北京卫平智业专利代理事务所(普通合伙) 11392

代理人 谢建玲 郝亮

(51)Int.Cl.

G06F 17/30(2006.01)

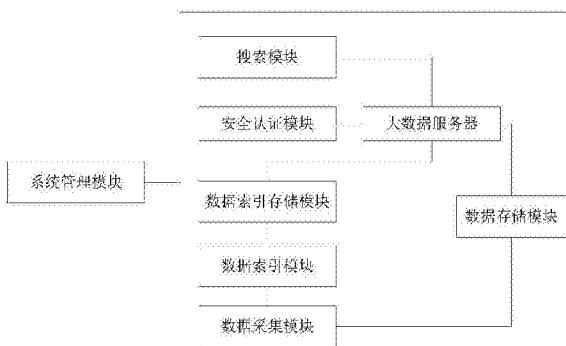
权利要求书2页 说明书4页 附图1页

(54)发明名称

一种基于大数据的搜索系统及方法

(57)摘要

本发明涉及一种基于大数据的搜索系统及方法,包括:系统管理模块、大数据服务器、安全认证模块、搜索模块、数据索引存储模块、数据采集模块、数据索引模块和数据存储模块;本发明所述的基于大数据的搜索系统,建立了专门的索引存储,并将历史搜索结果进行缓存利用,避免了重复搜索的工作量,提高了搜索的精准度和速度,可以从大量的数据中搜索到有价值的信息,很容易地解决了大数据时代对数据搜索的问题。



1. 一种基于大数据的搜索系统,其特征在于,包括:系统管理模块、大数据服务器、安全认证模块、搜索模块、数据索引存储模块、数据采集模块、数据索引模块和数据存储模块;

所述系统管理模块用于为其他模块提供接口服务;

所述搜索模块用于将用户的搜索请求提交给大数据服务器,并将搜索结果反馈给用户;

所述大数据服务器用于接收搜索模块的搜索请求,并从数据索引存储模块中进行搜索,然后将搜索结果与数据存储模块中的数据进行匹配,最后将匹配好的结果返回给搜索模块;

所述安全认证模块用于对大数据服务器接收的搜索请求进行安全认证;若通过认证,则将继续处理该搜索请求,若未通过认证,则向用户发送包含拒绝信息的搜索响应;

所述数据索引模块用于将数据采集模块采集到的数据信息按照数据属性对数据建立索引,并将建立的索引发送到数据索引存储模块进行存储;

所述数据索引存储模块用于存储数据索引模块建立的数据索引信息;

所述数据采集模块用于采集所需的数据,并对采集到的数据进行处理,处理成便于建立索引并搜索的数据格式;

所述数据存储模块用于存储数据采集模块采集到的各类数据。

2. 如权利要求1所述的基于大数据的搜索系统,其特征在于,所述搜索模块包括搜索优化模块,所述搜索优化模块用于获取用户的历史搜索结果,如搜索请求已有搜索结果,则不再响应用户提出的相同搜索请求,并将已有的搜索结果反馈给用户。

3. 如权利要求1所述的基于大数据的搜索系统,其特征在于,所述数据采集模块包括数据分解单元,用于将采集到的数据分解成至少一个关键数据。

4. 如权利要求1所述的基于大数据的搜索系统,其特征在于,所述数据采集模块采集的数据包括结构化数据和非结构化数据。

5. 如权利要求4所述的基于大数据的搜索系统,其特征在于,所述非结构化数据包括Word文档、Excel表格、PDF文件、图片、网页和视频。

6. 如权利要求1所述的基于大数据的搜索系统,其特征在于,所述数据属性包括类型、标题和日期。

7. 一种基于大数据的搜索方法,应用上述权利要求1-6任一权利要求所述的基于大数据的搜索系统,其特征在于,包括如下步骤:

(1) 利用数据采集模块采集数据;

(2) 将采集的数据存储至数据存储模块;

(3) 同时数据索引模块对采集的数据建立索引;

(4) 将建立的索引数据存储至数据索引存储模块;

(5) 用户通过搜索模块向大数据服务器提出搜索请求;

(6) 安全认证模块对搜索请求进行安全认证,如认证通过,则执行搜索过程,否则拒绝搜索,并反馈给用户;

(7) 通过安全认证后的搜索请求,如与搜索优化模块获取的历史搜索请求相同,则将相应的历史搜索结果返回给搜索模块,执行第(10)步,否则执行第(8)步;

(8) 大数据服务器接收搜索请求,从数据索引存储模块中进行搜索;

- (9) 大数据服务器将数据索引搜索结果与数据存储模块中的数据进行匹配,将匹配好的数据搜索结果返回给搜索模块;
- (10) 搜索模块将搜索结果反馈给用户。

一种基于大数据的搜索系统及方法

技术领域

[0001] 本发明涉及数据搜索技术领域,具体而言,涉及一种基于大数据的搜索系统及方法。

背景技术

[0002] 随着计算机和网络的发展,搜索功能已经成为最常用的功能,用户通过搜索操作方便快捷地获取目的信息。但是,随着业务的发展,可搜索的数据量也越来越大,目前将不断增大的数据量称之为大数据,对其的搜索操作称之为对大数据的搜索。

[0003] 大数据的影响渗透到了各个行业,对未来的科技和经济发展带来了深远影响,然而,大数据带来便利的同时,越来越多的人类活动每个月都会产生PB级的数据,同时越来越多的需求也需要每天处理几十PB、几百PB甚至EB级的数据,由于其数据量极其巨大,通过搜索系统,需要要处理的信息量越来越大,导致出现搜索效率变低,极大地影响了搜索效率,增加了大数据中心的工作负荷、时间消耗和设备损耗。

[0004] 虽然现有已经有各种各样的数据搜索系统,但是面对如此庞大的数据,如何从海量数据中快速获取需要的数据,并挖掘需要的知识,是当今面临的一个挑战。

发明内容

[0005] 针对现有技术中存在的缺陷,本发明的目的在于提供一种基于大数据的搜索系统,有助于提升搜索速度,可以从大量的数据中搜索到有价值的信息。

[0006] 为达到以上目的,本发明采取的技术方案是:

[0007] 一种基于大数据的搜索系统,包括:系统管理模块、大数据服务器、安全认证模块、搜索模块、数据索引存储模块、数据采集模块、数据索引模块和数据存储模块;

[0008] 所述系统管理模块用于为其他模块提供接口服务;

[0009] 所述搜索模块用于将用户的搜索请求提交给大数据服务器,并将搜索结果反馈给用户;

[0010] 所述大数据服务器用于接收搜索模块的搜索请求,并从数据索引存储模块中进行搜索,然后将搜索结果与数据存储模块中的数据进行匹配,最后将匹配好的结果返回给搜索模块;

[0011] 所述安全认证模块用于对大数据服务器接收的搜索请求进行安全认证;若通过认证,则将继续处理该搜索请求,若未通过认证,则向用户发送包含拒绝信息的搜索响应;

[0012] 所述数据索引模块用于将数据采集模块采集到的数据信息按照数据属性对数据建立索引,并将建立的索引发送到数据索引存储模块进行存储;

[0013] 所述数据索引存储模块用于存储数据索引模块建立的数据索引信息;

[0014] 所述数据采集模块用于采集所需的数据,并对采集到的数据进行处理,处理成便于建立索引并搜索的数据格式;

[0015] 所述数据存储模块用于存储数据采集模块采集到的各类数据。

[0016] 在上述方案的基础上,所述搜索模块包括搜索优化模块,所述搜索优化模块用于获取用户的历史搜索结果,如搜索请求已有搜索结果,则不再响应用户提出的相同搜索请求,并将已有的搜索结果反馈给用户。

[0017] 在上述方案的基础上,所述数据采集模块包括数据分解单元,用于将采集到的数据分解成至少一个关键数据。

[0018] 在上述方案的基础上,所述数据采集模块采集的数据包括结构化数据和非结构化数据。

[0019] 在上述方案的基础上,所述非结构化数据包括Word文档、Excel表格、PDF文件、图片、网页和视频。

[0020] 在上述方案的基础上,所述数据属性包括类型、标题和日期。

[0021] 一种基于大数据的搜索方法,应用上述基于大数据的搜索系统,包括如下步骤:

[0022] (1)利用数据采集模块采集数据;

[0023] (2)将采集的数据存储至数据存储模块;

[0024] (3)同时数据索引模块对采集的数据建立索引;

[0025] (4)将建立的索引数据存储至数据索引存储模块;

[0026] (5)用户通过搜索模块向大数据服务器提出搜索请求;

[0027] (6)安全认证模块对搜索请求进行安全认证,如认证通过,则执行搜索过程,否则拒绝搜索,并反馈给用户;

[0028] (7)通过安全认证后的搜索请求,如与搜索优化模块获取的历史搜索请求相同,则将相应的历史搜索结果返回给搜索模块,执行第(10)步,否则执行第(8)步;

[0029] (8)大数据服务器接收搜索请求,从数据索引存储模块中进行搜索;

[0030] (9)大数据服务器将数据索引搜索结果与数据存储模块中的数据进行匹配,将匹配好的数据搜索结果返回给搜索模块;

[0031] (10)搜索模块将搜索结果反馈给用户。

[0032] 本发明所述的基于大数据的搜索系统及方法,建立了专门的索引存储,并将历史搜索结果进行缓存利用,避免了重复搜索的工作量,提高了搜索的精准度和速度,可以从大量的数据中搜索到有价值的信息,很容易地解决了大数据时代对数据搜索的问题。

附图说明

[0033] 本发明有如下附图:

[0034] 图1本发明所述的基于大数据的搜索系统结构示意图。

[0035] 图2本发明所述的基于大数据的搜索方法流程示意图。

具体实施方式

[0036] 以下结合附图对本发明作进一步详细说明。

[0037] 如图1所示,本发明所述的基于大数据的搜索系统,包括:系统管理模块、大数据服务器、安全认证模块、搜索模块、数据索引存储模块、数据采集模块、数据索引模块和数据存储模块;

[0038] 所述系统管理模块用于为其他模块提供接口服务;

[0039] 所述搜索模块用于将用户的搜索请求提交给大数据服务器，并将搜索结果反馈给用户；

[0040] 所述大数据服务器用于接收搜索模块的搜索请求，并从数据索引存储模块中进行搜索，然后将搜索结果与数据存储模块中的数据进行匹配，最后将匹配好的结果返回给搜索模块；

[0041] 所述安全认证模块用于对大数据服务器接收的搜索请求进行安全认证；若通过认证，则将继续处理该搜索请求，若未通过认证，则向用户发送包含拒绝信息的搜索响应；

[0042] 所述数据索引模块用于将数据采集模块采集到的数据信息按照数据属性对数据建立索引，并将建立的索引发送到数据索引存储模块进行存储；

[0043] 所述数据索引存储模块用于存储数据索引模块建立的数据索引信息；

[0044] 所述数据采集模块用于采集所需的数据，并对采集到的数据进行处理，处理成便于建立索引并搜索的数据格式；

[0045] 所述数据存储模块用于存储数据采集模块采集到的各类数据。

[0046] 在上述方案的基础上，所述搜索模块包括搜索优化模块，所述搜索优化模块用于获取用户的历史搜索结果，如搜索请求已有搜索结果，则不再响应用户提出的相同搜索请求，并将已有的搜索结果反馈给用户。

[0047] 在上述方案的基础上，所述数据采集模块包括数据分解单元，用于将采集到的数据分解成至少一个关键数据。

[0048] 在上述方案的基础上，所述数据采集模块采集的数据包括结构化数据和非结构化数据。

[0049] 在上述方案的基础上，所述非结构化数据包括Word文档、Excel表格、PDF文件、图片、网页和视频。

[0050] 在上述方案的基础上，所述数据属性包括类型、标题和日期。

[0051] 如图2所示，一种基于大数据的搜索方法，应用上述基于大数据的搜索系统，包括如下步骤：

[0052] (1) 利用数据采集模块采集数据；

[0053] (2) 将采集的数据存储至数据存储模块；

[0054] (3) 同时数据索引模块对采集的数据建立索引；

[0055] (4) 将建立的索引数据存储至数据索引存储模块；

[0056] (5) 用户通过搜索模块向大数据服务器提出搜索请求；

[0057] (6) 安全认证模块对搜索请求进行安全认证，如认证通过，则执行搜索过程，否则拒绝搜索，并反馈给用户；

[0058] (7) 通过安全认证后的搜索请求，如与搜索优化模块获取的历史搜索请求相同，则将相应的历史搜索结果返回给搜索模块，执行第(10)步，否则执行第(8)步；

[0059] (8) 大数据服务器接收搜索请求，从数据索引存储模块中进行搜索；

[0060] (9) 大数据服务器将数据索引搜索结果与数据存储模块中的数据进行匹配，将匹配好的数据搜索结果返回给搜索模块；

[0061] (10) 搜索模块将搜索结果反馈给用户。

[0062] 本发明所述的基于大数据的搜索系统及方法，提高了搜索的精准度和速度，可以

从大量的数据中搜索到有价值的信息，解决了大数据时代对数据搜索的问题。

[0063] 本说明书中未做详细描述的内容属于本领域专业技术人员公知的现有技术。

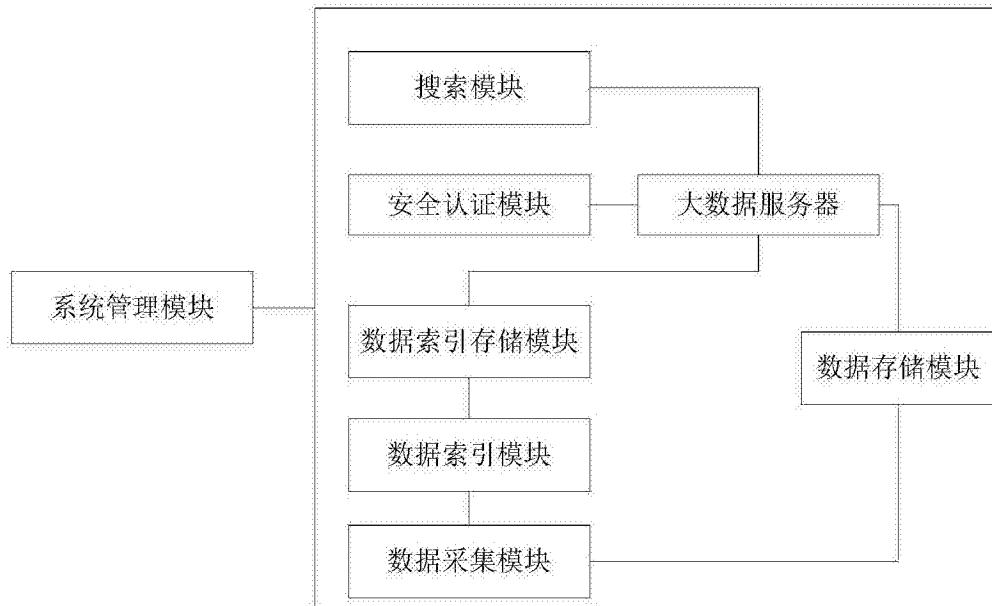


图1

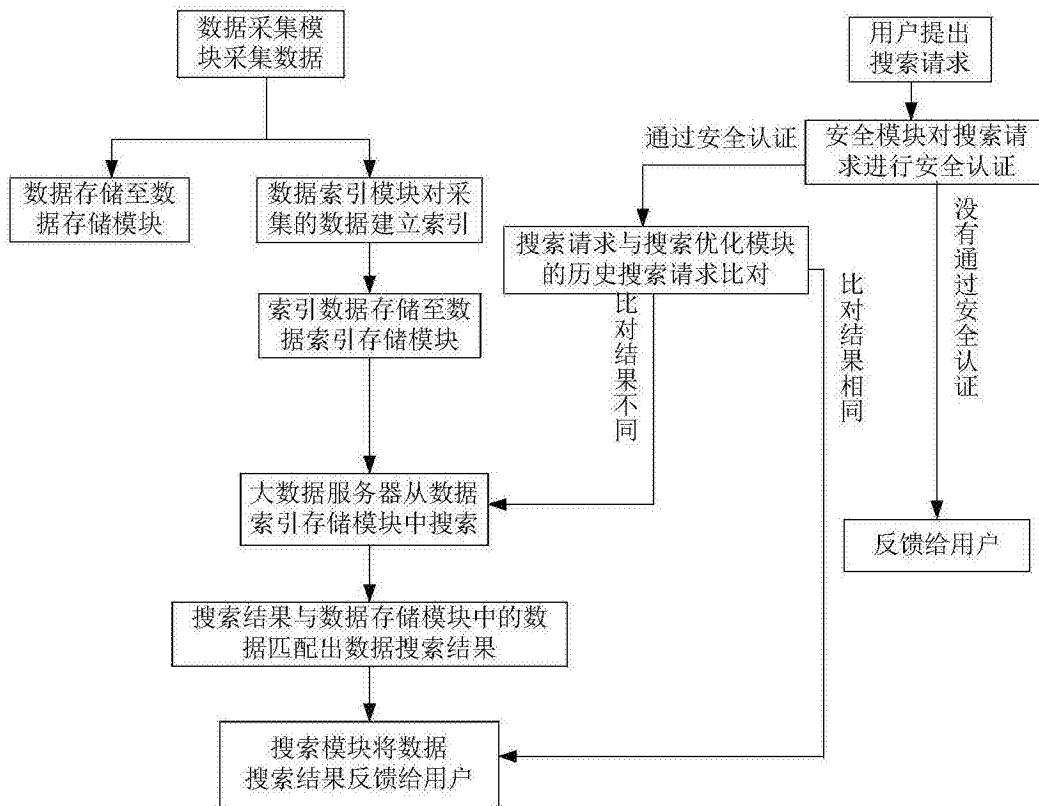


图2