(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105653981 B
(45)授权公告日 2018.11.30

(21)申请号 201511026582.1

(22)申请日 2015.12.31

(65)同一申请的已公布的文献号
申请公布号 CN 105653981 A

(43)申请公布日 2016.06.08

(73)专利权人 中国电子科技网络信息安全有限公司
地址 610041 四川省成都市双流县西南航空港经济开发区工业集中区内

(72)发明人 陈天莹 李全兵 李霄

(51)Int.Cl.
G06F 21/62(2013.01)

(56)对比文件
US 2015/0278545 A1,2015.10.01,
CN 104866775 A,2015.08.26,
CN 105022832 A,2015.11.04,
刘明辉等.云环境下的敏感数据保护技术研究.《电信科学》.2014,(第11期),第3-4页及图1.
闫玺玺.开放网络环境下敏感数据安全与防泄密关键技术研究.《中国博士学位论文全文数据库 信息科技辑》.2013,(第01期),说明书第28-30页.

审查员 周辉

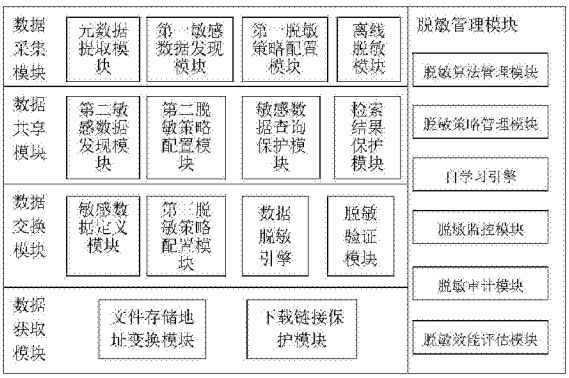
权利要求书5页 说明书11页 附图6页

(54)发明名称

大数据平台的数据流通与交易的敏感数据保护系统及方法

(57)摘要

一种大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述大数据平台的数据流通与交易的敏感数据保护系统包括在数据采集中发现敏感内容并对敏感内容进行保护处理的数据采集模块、对数据共享过程中的敏感数据进行保护处理的数据共享模块、对数据交换过程中的相对敏感数据配置脱敏策略进行脱敏处理的数据交换模块、在数据获取过程中对数据文件下载链接及存储地址进行保护的数据获取模块、对敏感数据的脱敏及保护处理进行管理和监控以及审计的脱敏管理模块。本发明还公开了一种大数据平台的数据流通与交易的敏感数据保护方法。



CN 105653981 B

1.一种大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述大数据平台的数据流通与交易的敏感数据保护系统包括在数据采集中发现敏感内容并对敏感内容进行保护处理的数据采集模块、对数据共享过程中的敏感数据进行保护处理的数据共享模块、对数据交换过程中的相对敏感数据配置脱敏策略进行脱敏处理的数据交换模块、在数据获取过程中对数据文件下载链接及存储地址进行保护的数据获取模块、对敏感数据的脱敏及保护处理进行管理和监控以及审计的脱敏管理模块；所述数据采集模块包括对上传大数据平台的数据进行数据信息提取为敏感数据保护提供数据准备的元数据提取模块、在所述元数据提取模块提取的数据信息基础上自动发现涉密信息及敏感数据的第一敏感数据发现模块、为所述第一敏感数据发现模块发现的敏感内容配置相应的脱敏算法形成脱敏策略的第一脱敏策略配置模块、通过系统调用所述第一脱敏策略配置模块预定义的脱敏策略对敏感数据实现批量离线脱敏的离线脱敏模块。

2.根据权利要求1所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述元数据提取模块提取上传大数据平台的数据的数据背景、数据内容、数据结构、存储位置信息；所述第一敏感数据发现模块通过设定敏感内容的检查范围、敏感内容的背景信息，采用基于规则和数据挖掘的方法自动发现数据中的敏感内容；所述第一敏感内容脱敏策略配置模块根据所述第一敏感数据发现模块发现的敏感内容的属性不同配置相应的脱敏算法形成相应的脱敏策略并同时按照敏感内容属性预定义脱敏策略。

3.根据权利要求1所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述数据共享模块包括对存储于大数据平台中的数据根据数据属性选择采用人工定义和自动发现方式中的一种进行敏感数据发现的第二敏感数据发现模块、在所述第二敏感数据发现模块发现的敏感数据基础上为每一类敏感数据配置脱敏算法形成脱敏策略的第二脱敏策略配置模块、对大数据平台中允许共享的数据进行噪声干扰处理保护敏感数据的敏感数据查询保护模块、对大数据平台中的数据检索结果进行数据脱敏保护的检索结果保护模块。

4.根据权利要求3所述的大数据平台的数据流通与交易的敏感数据保护系统，所述第二敏感数据发现模块采用的人工定义方式发现敏感数据是由资源发布人依据个人经验定义敏感数据，所述自动发现方式是基于专家系统和自然语言处理方式对敏感数据进行自动发现并为资源发布人推荐敏感数据；所述第二脱敏策略配置模块根据敏感数据的特点推荐脱敏算法形成脱敏策略或者自行定制脱敏算法形成新的脱敏策略，并对已形成的脱敏策略进行存储和使用率统计分析以实现后续脱敏策略自动推荐预定义；所述敏感数据查询保护模块对数据需求方在大数据平台中的数据查询结果通过对原始数据、原始数据的转换、统计结果使用拉普拉斯机制和指数机制实现差分隐私添加噪音来达到保护敏感数据的目的；所述检索结果保护模块对大数据平台资源发布人允许共享的数据的检索结果中的敏感信息采用遮挡、置换的方式进行脱敏处理。

5.根据权利要求3所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述数据交换模块包括针对数据需求方进行敏感数据定义的敏感数据定义模块、为所述敏感数据定义模块定义的敏感数据配置相应的脱敏策略的第三脱敏策略配置模块、根据所述第三脱敏策略配置模块配置的脱敏策略对数据执行脱敏处理的数据脱敏引擎、对脱敏结果的正确性和真实性进行验证的脱敏验证模块。

6.根据权利要求5所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，敏感数据定义模块由资源发布人根据已经定义的极敏感数据信息和数据需求方的身份、数据使用权限，修改原先预定义的敏感数据，定义针对于数据需求方的敏感数据；所述第三脱敏策略配置模块确认由所述第二脱敏策略配置模块为所述敏感数据定义模块定义的针对于数据需求方的敏感数据配置的脱敏策略并在脱敏策略不合适时进行形成最终脱敏策略；所述数据脱敏引擎根据所述第三脱敏策略模块确定的最终脱敏策略对所述敏感数据定义模块定义的敏感数据进行脱敏处理；所述脱敏验证模块将脱敏前数据与脱敏后数据进行数据格式、长度和完整性的检查，并使用相应脱敏规则对数据进行脱敏得到脱敏结果与所述脱敏引擎脱敏后的数据进行比对验证脱敏的准确性，同时对原有数据的逻辑关系和统计分布进行比对，验证数据脱敏的真实性。

7.根据权利要求1所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述数据获取模块包括对文件存储的真实地址进行变换形成新的存储地址达到存储地址保护的文件存储地址变换模块、在所述文件存储地址编号模块对文件真实存储地址进行变换的基础上为有下载需求的数据需求方展示变换后的文件存储地址防止文件存储地址泄露保护下载链接安全的下载链接保护模块。

8.根据权利要求7所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述文件存储地址变换模块采用散列函数对文件的原始地址进行计算生成新的存储地址。

9.根据权利要求2至8任一所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述脱敏管理模块包括对数据脱敏算法进行添加、删除和修改的脱敏算法管理模块、对数据流通过程中的数据脱敏策略进行增加、删除和修改的脱敏策略管理模块、对敏感内容和敏感数据特征分析的基础上训练得到机器学习模型并在有敏感数据自动发现需求时对敏感数据进行自动发现的自学习引擎、对数据流通过程中的各个环节的数据脱敏进行实时监控以便及时发现异常的脱敏监控模块、对数据流通和交易过程中的数据脱敏任务进行分析审计的脱敏审计模块、对数据流通和交易过程中敏感数据保护处理进行评价的脱敏效能评估模块。

10.根据权利要求9所述的大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述脱敏算法管理模块向系统添加新的脱敏算法及相应的算法描述并可删除不适用的脱敏算法；所述脱敏策略管理模块对系统中的脱敏策略的使用情况进行挖掘分析添加用户常用的脱敏策略，修改不合适的脱敏策略并删除无用脱敏策略；所述脱敏监控模块对数据流通和交易中敏感数据的脱敏状态、脱敏策略、脱敏结果、数据需求方进行关联分析和挖局，及时发现数据脱敏过程中的异常，以便在发生数据泄露是能实现数据追溯；所述脱敏效能评估模块对数据流通和交易中敏感数据脱敏保护的功能、敏感数据脱敏保护的应用场景、敏感数据脱敏保护的正确性和真实性以及有效性进行综合评价。

11.一种大数据平台的数据流通和交易的敏感数据保护方法，其特征在于，所述大数据平台的数据流通和交易的敏感数据保护方法采用大数据平台的数据流通和交易的敏感数据保护系统来实现，包括如下步骤：

步骤一、资源发布人在数据采集前发现数据中的敏感内容，对数据采集中的敏感数据进行保护处理再将数据上传到大数据平台；

步骤二、数据上传到大数据平台后，资源发布人对数据进行共享时，对数据共享中的敏感数据进行保护之后再共享发布；

步骤三、数据需求方对资源发布人共享的数据请求交换，资源发布人对数据需求方请求交换的数据中的敏感内容进行定义，对数据交换中的敏感数据进行保护处理后再交换给数据需求方；

步骤四、数据需求方请求下载资源发布人共享的数据，对数据下载中的敏感数据存储地址进行保护；

所述对数据采集中的敏感数据进行保护包括如下步骤：

s11、资源发布人在向大数据平台上传数据前，制定元数据提取标准；进入步骤s12；

s12、解析上传的数据，进入步骤s13；

s13、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的元数据提取模块根据所述步骤s11中制定的元数据提取标准提取经步骤s12解析后的上传数据的元数据信息，进入步骤s14；

s14、根据所述步骤s13中提取的元数据信息进行敏感内容配置，进入步骤s15；

s15通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第一敏感数据发现模块根据所述步骤s14中配置的敏感内容，对数据中的敏感内容进行自动识别，进入步骤s16；

s16、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第一脱敏策略配置模块根据所述步骤s15中识别确定的敏感内容的特点，自动配置脱敏算法，形成脱敏策略，进入步骤s17；

s17、根据所述步骤s16中配置的脱敏算法，对数据进行脱敏处理，进入步骤s18；

s18、将脱敏后的数据上传到大数据平台。

12.根据权利要求11所述的大数据平台的数据流通和交易的敏感数据保护方法，其特征在于，所述对数据共享中的敏感数据进行保护包括如下步骤：

s21、资源发布人上传数据到大数据平台，进入步骤s22；

s22、在大数据平台环境下，通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第二敏感数据发现模块根据数据属性定义敏感数据，进入步骤s23；

s23、根据所述步骤s22中定义的敏感数据确定数据中的敏感数据，进入步骤s24；

s24、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第二脱敏策略配置模块为所述步骤s23中确定的敏感数据配置脱敏算法，进入步骤s25；

s25、判断为敏感数据配置的脱敏算法是否合适，若否，进入步骤s26，若是，进入步骤s27；

s26、为敏感数据重新配置脱敏算法，进入步骤s27；

s27、根据配置的脱敏算法形成脱敏策略，进入步骤s28；

s28、保存脱敏策略，并对不同脱敏策略的使用情况进行统计分析，进入步骤s29；

s29、根据步骤s28中对脱敏策略使用情况的统计分析结果对脱敏策略使用率进行排序，进入步骤s210；

s210、根据脱敏策略使用率建立脱敏策略自动推荐机制。

13.根据权利要求12所述的大数据平台的数据流通和交易的敏感数据保护方法，其特

征在于，所述对数据交换中的敏感数据进行保护包括如下步骤：

s31、数据需求方输入查询条件对数据进行查询，进入步骤s32；

s32、对与查询条件匹配的数据进行差分隐私保护处理，进步步骤s33；

s33、将经过差分隐私保护处理的数据反馈给数据需求方，进步步骤s34；

s34、将反馈给数据需求方的数据以脱敏方式将数据显示给数据需求方供其查看，进入步骤s35；

s35、数据需求方请求共享数据，进入步骤s36；

s36、资源发布人审核数据需求方的身份和数据使用权限，以此为基础通过所述大数据平台的数据流通与交易的敏感数据保护系统中的敏感数据定义模块预定义相对于数据需求方的敏感数据，进入步骤s37；

s37、判断预定义的敏感数据与数据需求方身份与权限是否相符，若否，进入步骤s38，若是，进入步骤s39；

s38、重新定义敏感数据，进入步骤s39；

s39、根据定义确定相对于数据需求方的敏感数据，进入步骤s310；

s310、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第三脱敏策略配置模块判断确定的敏感数据的脱敏策略是否合适，若否，进入步骤s311，若是，进入步骤s312；

s311、修改脱敏策略，进入步骤s312；

s312、确认敏感数据脱敏策略，进入步骤s313；

s313、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的数据脱敏引擎根据确认的脱敏策略对数据需求方请求共享的数据进行脱敏处理，进入步骤s314；

s314、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的脱敏验证模块校验脱敏数据的正确性，进入步骤s315；

s315、向数据需求方展示校验后的可供共享的脱敏数据。

14.根据权利要求13所述的大数据平台的数据流通和交易的敏感数据保护方法，其特征在于，所述对数据下载中的敏感数据存储地址进行保护包括如下步骤：

S41、数据需求方从展示的结果中选择需要下载的数据文件，进入步骤s42；

S42、对数据需求方选择下载的文件进行下载链接保护处理，生成新的可映射到文件原存储地址的安全链接，进入步骤s43；

s43、向数据需求方展示新生成的安全链接供其下载文件。

15.根据权利要求14所述的大数据平台的数据流通和交易的敏感数据保护方法，其特征在于，通过所述大数据平台的数据流通与交易的敏感数据保护系统的脱敏管理模块对每一环节的脱敏算法、脱敏策略进行管理；

通过所述脱敏管理模块中的脱敏算法管理模块以实现为敏感数据配置合适脱敏算法的目的对每一环节的脱敏算法配置进行添加、删除和修改处理；

通过所述脱敏管理模块中的脱敏策略管理模块以实现为敏感数据配置合适脱敏策略的目的对数据流通和交易中的脱敏策略进行增加、删除和修改管理；

通过所述脱敏管理模块中的自学习引擎对数据流通和交易中的敏感内容和敏感数据特征进行分析训练以得到机器学习模型满足敏感数据的自动发现需求；

通过所述脱敏管理模块中的脱敏监控模块对数据流通中的各个环节的数据脱敏进行实时监控以便及时发现异常情况；

通过所述脱敏管理模块中的脱敏审计模块对数据流通和交易中的数据脱敏任务进行分析审计；

通过所述脱敏管理模块中的脱敏效能评估模块对数据流通和交易中的敏感数据的保护处理进行评价。

大数据平台的数据流通与交易的敏感数据保护系统及方法

技术领域

[0001] 本发明涉及大数据领域，具体地说，涉及大数据平台的数据流通与交易的敏感数据保护系统及方法。

背景技术

[0002] 在大数据环境下，数据的汇集、流通、交换共享、交易、分析挖掘等需求越来越强，大量的敏感数据汇集到大数据中心平台，如何在流通、交换共享、交易、分析挖掘等数据使用中保护这些敏感数据，防止用户隐私泄露已成为大数据安全关注的重点。目前，已有的敏感数据保护采用以下四种保护方式：

[0003] (1)访问控制的方法：通过对用户身份及其所属的安全等级来限制用户对数据的访问，防止敏感数据的未授权访问，实现敏感数据的保护。该方法在一定程度上降低了数据共享的可用性。

[0004] (2)基于数据失真的敏感数据保护技术：采用扰动、置换、遮挡等方法对敏感数据进行处理，在处理的同时保证数据保留某些统计的特征，以便进行数据分析与挖掘。该方法效率比较高，但是会使数据中的信息丢失。

[0005] (3)基于数据加密的技术：采用安全多方计算等加密技术对数据进行加密处理，保障数据流通过程中的敏感数据安全。该方法能有效地保障数据流通中数据的准确性和安全性，但是计算的效率相对较低，开销比较大。

[0006] (4)基于限制发布的技术：依据实际需求，借助差分隐私、k-匿名等算法对发布的数据进行处理，比如不发布数据中的某些属性或者对某些数据型进行泛化，实现敏感数据的保护。该方法能保证共享数据的真实性，但是会存在数据中信息的丢失。

[0007] 综上所述，存在的问题：

[0008] (1)已有的敏感数据保护方法大都集中在对数据流通的共享与发布环节，缺乏对数据流通整个环节的敏感数据的保护。

[0009] (2)过度关注用户隐私也会阻碍数据流通，如何结合不同敏感数据保护方法的优缺点为数据流通的不同环节选择恰当的技术，实现敏感数据保护与数据可用性之间的平衡，是亟待解决的问题。

[0010] (3)关于敏感数据保护技术的大都是直接对敏感数据处理，但是并未考虑到如何发现敏感数据、如何验证脱敏结果的正确性、真实性。

[0011] 因此提供一种大数据平台上数据流通过程中的敏感数据保护机制，保障数据流通过程中的敏感数据安全，平衡隐私保护与数据流通之间的关系，是迫切需要的。

发明内容

[0012] 为了达到上述目的，本发明提供一种保障数据流通过程中的数据采集、数据共享、数据交换以及数据获取整个过程敏感数据安全的大数据平台的数据流通与交易的敏感数据保护系统及方法。

[0013] 本发明的一种大数据平台的数据流通与交易的敏感数据保护系统，其特征在于，所述大数据平台的数据流通与交易的敏感数据保护系统包括在数据采集中发现敏感内容并对敏感内容进行保护处理的数据采集模块、对数据共享过程中的敏感数据进行保护处理的数据共享模块、对数据交换过程中的相对敏感数据配置脱敏策略进行脱敏处理的数据交换模块、在数据获取过程中对数据文件下载链接及存储地址进行保护的数据获取模块、对敏感数据的脱敏及保护处理进行管理和监控以及审计的脱敏管理模块。

[0014] 其中，所述数据采集模块包括对上传大数据平台的数据进行数据信息提取为敏感数据保护提供数据准备的元数据提取模块、在所述元数据提取模块提取的数据信息基础上自动发现涉密信息及敏感数据的第一敏感数据发现模块、为所述第一敏感数据发现模块发现的敏感内容配置相应的脱敏算法形成脱敏策略的第一脱敏策略配置模块、通过系统调用所述第一脱敏策略配置模块预定义的脱敏策略对敏感数据实现批量离线脱敏的离线脱敏模块。

[0015] 所述数据共享模块包括对存储于大数据平台中的数据根据数据属性选择采用人工定义和自动发现方式中的一种进行敏感数据发现的第二敏感数据发现模块、在所述第二敏感数据发现模块发现的敏感数据基础上为每一类敏感数据配置脱敏算法形成脱敏策略的第二脱敏策略配置模块、对大数据平台中允许共享的数据进行噪声干扰处理保护敏感数据的敏感数据查询保护模块、对大数据平台中的数据检索结果进行数据脱敏保护的检索结果保护模块。

[0016] 所述数据获取模块包括对文件存储的真实地址进行变换形成新的存储地址达到存储地址保护的文件存储地址变换模块、在所述文件存储地址编号模块对文件真实存储地址进行变换的基础上为有下载需求的数据需求方展示变换后的文件存储地址防止文件存储地址泄露保护下载链接安全的下载链接保护模块。

[0017] 所述脱敏管理模块包括对数据脱敏算法进行添加、删除和修改的脱敏算法管理模块、对数据流通过过程中的数据脱敏策略进行增加、删除和修改的脱敏策略管理模块、对敏感内容和敏感数据特征分析的基础上训练得到机器学习模型并在有敏感数据自动发现需求时对敏感数据进行自动发现的自学习引擎、对数据流通过程中的各个环节的数据脱敏进行实时监控以便及时发现异常的脱敏监控模块、对数据流通和交易过程中的数据脱敏任务进行分析审计的脱敏审计模块、对数据流通和交易过程中敏感数据保护处理进行评价的脱敏效能评估模块。

[0018] 本发明的一种大数据平台的数据流通和交易的敏感数据保护方法，采用大数据平台的数据流通和交易的敏感数据保护系统来实现，包括如下步骤：

[0019] 步骤一、资源发布人在数据采集前发现数据中的敏感内容，对数据采集中的敏感数据进行保护处理再将数据上传到大数据平台；

[0020] 步骤二、数据上传到大数据平台后，资源发布人对数据进行共享时，对数据共享中的敏感数据进行保护之后再共享发布；

[0021] 步骤三、数据需求方对资源发布人共享的数据请求交换，资源发布人对数据需求方请求交换的数据中的敏感内容进行定义，对数据交换中的敏感数据进行保护处理后再交换给数据需求方；

[0022] 步骤四、数据需求方请求下载资源发布人共享的数据，对数据下载中的敏感数据

存储地址进行保护。

[0023] 其中，所述对数据采集中的敏感数据进行保护包括如下步骤：

[0024] s11、资源发布人在向大数据平台上传数据前，制定元数据提取标准；进入步骤s12；

[0025] s12、解析上传的数据，进入步骤s13；

[0026] s13、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的元数据提取模块根据所述步骤s11中制定的元数据提取标准提取经步骤s12解析后的上传数据的元数据信息，进入步骤s14；

[0027] s14、根据所述步骤s13中提取的元数据信息进行敏感内容配置，进入步骤s15；

[0028] s15通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第一敏感数据发现模块根据所述步骤s14中配置的敏感内容，对数据中的敏感内容进行自动识别，进入步骤s16；

[0029] s16、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第一脱敏策略配置模块根据所述步骤s15中识别确定的敏感内容的特点，自动配置脱敏算法，形成脱敏策略，进入步骤s17；

[0030] s17、根据所述步骤s16中配置的脱敏算法，对数据进行脱敏处理，进入步骤s18；

[0031] s18、将脱敏后的数据上传到大数据平台。

[0032] 所述对数据共享中的敏感数据进行保护包括如下步骤：

[0033] s21、资源发布人上传数据到大数据平台，进入步骤s22；

[0034] s22、在大数据平台环境下，通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第二敏感数据发现模块根据数据属性定义敏感数据，进入步骤s23；

[0035] s23、根据所述步骤s22中定义的敏感数据确定数据中的敏感数据，进入步骤s24；

[0036] s24、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第二脱敏策略配置模块为所述步骤s23中确定的敏感数据配置脱敏算法，进入步骤s25；

[0037] s25、判断为敏感数据配置的脱敏算法是否合适，若否，进入步骤s26，若是，进入步骤s27；

[0038] s26、为敏感数据重新配置脱敏算法，进入步骤s27；

[0039] s27、根据配置的脱敏算法形成脱敏策略，进入步骤s28；

[0040] s28、保存脱敏策略，并对不同脱敏策略的使用情况进行统计分析，进入步骤s29；

[0041] s29、根据步骤s28中对脱敏策略使用情况的统计分析结果对脱敏策略使用率进行排序，进入步骤s210；

[0042] s210、根据脱敏策略使用率建立脱敏策略自动推荐机制。

[0043] 所述对数据交换中的敏感数据进行保护包括如下步骤：

[0044] s31、数据需求方输入查询条件对数据进行查询，进入步骤s32；

[0045] s32、对与查询条件匹配的数据进行差分隐私保护处理，进步步骤s33；

[0046] s33、将经过差分隐私保护处理的数据反馈给数据需求方，进步步骤s34；

[0047] s34、将反馈给数据需求方的数据以脱敏方式将数据显示给数据需求方供其查看，进入步骤s35；

[0048] s35、数据需求方请求共享数据，进入步骤s36；

[0049] s36、资源发布人审核数据需求方的身份和数据使用权限，以此为基础通过所述大数据平台的数据流通与交易的敏感数据保护系统中的敏感数据定义模块预定义相对于数据需求方的敏感数据，进入步骤s37；

[0050] s37、判断预定义的敏感数据与数据需求方身份与权限是否相符，若否，进入步骤s38，若是，进入步骤s39；

[0051] s38、重新定义敏感数据，进入步骤s39；

[0052] s39、根据定义确定相对于数据需求方的敏感数据，进入步骤s310；

[0053] s310、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第三脱敏策略配置模块判断确定的敏感数据的脱敏策略是否合适，若否，进入步骤s311，若是，进入步骤s312；

[0054] s311、修改脱敏策略，进入步骤s312；

[0055] s312、确认敏感数据脱敏策略，进入步骤s313；

[0056] s313、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的数据脱敏引擎根据确认的脱敏策略对数据需求方请求共享的数据进行脱敏处理，进入步骤s314；

[0057] s314、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的脱敏验证模块校验脱敏数据的正确性，进入步骤s315；

[0058] s315、向数据需求方展示校验后的可供共享的脱敏数据。

[0059] 所述对数据下载中的敏感数据存储地址进行保护包括如下步骤：

[0060] S41、数据需求方从展示的结果中选择需要下载的数据文件，进入步骤s42；

[0061] S42、对数据需求方选择下载的文件进行下载链接保护处理，生成新的可映射到文件原存储地址的安全链接，进入步骤s43；

[0062] s43、向数据需求方展示新生成的安全链接供其下载文件。

[0063] 本发明的有益效果在于：(1)从数据流通的整个环节实现了敏感数据的保护；(2)在数据流通的不同环节使用不同的敏感数据保护方法，实现敏感数据保护与数据可用性之间的平衡；(3)提出了基于专家系统和自然语言处理的敏感数据自动发现方法，能够自动发现敏感数据；(4)提出了验证脱敏结果正确性与真实性的方法，能够有效地度量数据脱敏环节。

附图说明

[0064] 图1是本发明的大数据平台的数据流通与交易的敏感数据保护系统的框架结构示意图；

[0065] 图2是本发明的大数据平台的数据流通与交易的敏感数据保护方法的主体流程示意图；

[0066] 图3是本发明的大数据平台的数据流通与交易的敏感数据保护方法的数据采集的敏感数据保护流程示意图；

[0067] 图4是本发明的大数据平台的数据流通与交易的敏感数据保护方法的数据共享的敏感数据确定流程示意图；

[0068] 图5是本发明的大数据平台的数据流通与交易的敏感数据保护方法的数据交换的敏感数据查询保护流程示意图；

[0069] 图6是本发明的大数据平台的数据流通与交易的敏感数据保护方法的数据交换的敏感数据查询结果交换保护流程示意图；

[0070] 图7是本发明的大数据平台的数据流通与交易的敏感数据保护方法的数据交换的敏感数据保护流程示意图。

具体实施方式

[0071] 为了更好的理解本发明，下面结合附图详细说明本发明。

[0072] 如图1所示，本发明的一种大数据平台的数据流通与交易的敏感数据保护系统，包括在数据采集中发现敏感内容并对敏感内容进行保护处理的数据采集模块、对数据共享过程中的敏感数据进行保护处理的数据共享模块、对数据交换过程中的相对敏感数据配置脱敏策略进行脱敏处理的数据交换模块、在数据获取过程中对数据文件下载链接及存储地址进行保护的数据获取模块、对敏感数据的脱敏及保护处理进行管理和监控以及审计的脱敏管理模块。

[0073] 其中，所述数据采集模块包括对上传大数据平台的数据进行数据信息提取为敏感数据保护提供数据准备的元数据提取模块、在所述元数据提取模块提取的数据信息基础上自动发现涉密信息及敏感数据的第一敏感数据发现模块、为所述第一敏感数据发现模块发现的敏感内容配置相应的脱敏算法形成脱敏策略的第一脱敏策略配置模块、通过系统调用所述第一脱敏策略配置模块预定义的脱敏策略对敏感数据实现批量离线脱敏的离线脱敏模块。

[0074] 优选地，所述元数据提取模块提取上传大数据平台的数据的数据背景、数据内容、数据结构、存储位置信息；所述第一敏感数据发现模块通过设定敏感内容的检查范围、敏感内容的背景信息，采用基于规则和数据挖掘的方法自动发现数据中的敏感内容；所述第一敏感内容脱敏策略配置模块根据所述第一敏感数据发现模块发现的敏感内容的属性不同配置相应的脱敏算法形成相应的脱敏策略并同时按照敏感内容属性预定义脱敏策略。

[0075] 具体地说，资源发布人在数据采集前发现数据中的敏感内容，对敏感内容进行处理是防止敏感、涉密信息泄露的基础环节。元数据提取模块对待上传的文档、传统数据库与分布式数据库等数据的背景、内容、数据结构、存储位置等信息提取出来，为敏感数据保护提供数据准备。第一敏感数据发现模块在元数据提取的基础上，自动发现数据中的涉密及敏感信息。用户可通过对敏感内容、敏感内容的检查范围、敏感内容背景信息的设定，采用基于规则和数据挖掘的方法自动发现数据中的敏感内容。第一脱敏策略配置模块在敏感内容发现模块的基础上，为敏感内容配置相应的数据脱敏算法，形成脱敏策略。离线脱敏模块在本地对数据进行脱敏。系统调用敏感内容脱敏策略配置模块预定义的脱敏策略及敏感数据，实现敏感数据的批量脱敏。

[0076] 所述数据共享模块包括对存储于大数据平台中的数据根据数据属性选择采用人工定义和自动发现方式中的一种进行敏感数据发现的第二敏感数据发现模块、在所述第二敏感数据发现模块发现的敏感数据基础上为每一类敏感数据配置脱敏算法形成脱敏策略的第二脱敏策略配置模块、对大数据平台中允许共享的数据进行噪声干扰处理保护敏感数据的敏感数据查询保护模块、对大数据平台中的数据检索结果进行数据脱敏保护的检索结果保护模块。

[0077] 优选地，所述第二敏感数据发现模块采用的人工定义方式发现敏感数据是由资源发布人依据个人经验定义敏感数据，所述自动发现方式是基于专家系统和自然语言处理方式对敏感数据进行自动发现并为资源发布人推荐敏感数据；所述第二脱敏策略配置模块根据敏感数据的特点推荐脱敏算法形成脱敏策略或者自行定制脱敏算法形成新的脱敏策略，并对已形成的脱敏策略进行存储和使用率统计分析以实现后续脱敏策略自动推荐预定义；所述敏感数据查询保护模块对数据需求方在大数据平台中的数据查询结果通过对原始数据、原始数据的转换、统计结果使用拉普拉斯机制和指数机制实现差分隐私添加噪音来达到保护敏感数据的目的；所述检索结果保护模块对大数据平台资源发布人允许共享的数据的检索结果中的敏感信息采用遮挡、置换的方式进行脱敏处理。

[0078] 简单地说，数据共享是指数据需求方通过大数据平台对平台数据进行查询分析，知悉所需要的数据信息。数据共享模块是对平台数据共享过程中的敏感数据进行保护。第二敏感数据发现模块采用人工定义和自动发现两种方式进行敏感数据发现。当数据属性比较少时，采用人工定义敏感数据的方式，主要过程是资源发布人依据个人经验定义敏感数据；当数据属性较多时，采用自动发现的方式，主要是基于专家系统和自然语言处理两种方式对敏感数据进行自动发现，为资源发布人推荐敏感数据。第二脱敏策略配置模块是在敏感数据发现的基础上，为每一类敏感数据配置数据脱敏算法，形成脱敏策略。系统可依据数据的特点，为敏感数据推荐合适的算法，也可自行制定脱敏算法。当系统中策略达到一定的数量后，对策略的使用情况进行统计分析，实现策略的自动推荐。敏感数据查询分析保护模块是对平台数据的查询结果中添加适当噪音来达到敏感数据保护的结果。系统通过对原始数据、原始数据的准换、或者是对统计结果使用拉普拉斯机制和指数机制的方法实现差分隐私，添加噪音来达到敏感数据保护的结果。检索结果保护模块是对平台加密数据检索结果的数据脱敏保护。系统对检索的结果中数据来源、数据摘要、数据所有者等敏感信息，采用遮挡、置换等方式进行脱敏。

[0079] 所述数据交换模块包括针对数据需求方进行敏感数据定义的敏感数据定义模块、为所述敏感数据定义模块定义的敏感数据配置相应的脱敏策略的第三脱敏策略配置模块、根据所述第三脱敏策略配置模块配置的脱敏策略对数据执行脱敏处理的数据脱敏引擎、对脱敏结果的正确性和真实性进行验证的脱敏验证模块。

[0080] 优选地，敏感数据定义模块由资源发布人根据已经定义的极敏感数据信息和数据需求方的身份、数据使用权限，修改原先预定义的敏感数据，定义针对于数据需求方的敏感数据；所述第三脱敏策略配置模块确认由所述第二脱敏策略配置模块为所述敏感数据定义模块定义的针对于数据需求方的敏感数据配置的脱敏策略并在脱敏策略不合适时进行形成最终脱敏策略；所述数据脱敏引擎根据所述第三脱敏策略模块确定的最终脱敏策略对所述敏感数据定义模块定义的敏感数据进行脱敏处理；所述脱敏验证模块将脱敏前数据与脱敏后数据进行数据格式、长度和完整性的检查，并使用相应脱敏规则对数据进行脱敏得到脱敏结果与所述脱敏引擎脱敏后的数据进行比对验证脱敏的准确性，同时对原有数据的逻辑关系和统计分布进行比对，验证数据脱敏的真实性。

[0081] 数据交换是指数据需求方向资源发布人申请获取数据的过程。数据交换模块是资源发布人为数据需求方进行脱敏配置与脱敏处理。敏感数据定义模块是结合数据需求方的信息进行敏感数据的定义。资源发布人结合已定义的敏感数据信息和数据需求方的身份、

数据使用权限，修改敏感数据发现模块预定义的敏感数据。第三脱敏策略配置模块是在敏感数据定义的基础上，为敏感数据配置脱敏算法。系统在敏感数据定义模块的基础上，结合敏感数据的特征，对数据共享过程中预定义的脱敏策略进行确认与修改，如果敏感策略不合适进行修改，如果合适进行确认，形成最终的脱敏策略。数据脱敏引擎是对数据执行脱敏处理。当数据需求方发出数据交换请求时，平台调用预定义的敏感数据及脱敏策略，使用数据脱敏引擎对数据进行脱敏处理。脱敏验证模块是对数据脱敏的结果的正确性和真实性进行验证。系统将脱敏前数据与脱敏后数据进行数据格式、长度、完整性的检查，并使用相应脱敏规则对数据进行脱敏，将结果与脱敏后的数据进行比对，验证数据脱敏的正确性；对原有数据的逻辑关系与统计分布进行比对，验证数据脱敏的真实性。

[0082] 所述数据获取模块包括对文件存储的真实地址进行变换形成新的存储地址达到存储地址保护的文件存储地址变换模块、在所述文件存储地址编号模块对文件真实存储地址进行变换的基础上为有下载需求的数据需求方展示变换后的文件存储地址防止文件存储地址泄露保护下载链接安全的下载链接保护模块。优选地，所述文件存储地址变换模块采用散列函数对文件的原始地址进行计算生成新的存储地址。

[0083] 数据获取是数据需求方申请成功后下载数据的过程。数据获取模块是对数据需求方的数据下载过程进行保护。文件存储地址变换模块是对文件存储的真实地址进行变换，主要方法是采用散列函数对文件的原始地址进行计算，生成新的存储地址。下载链接保护模块是在文件存储地址变换模块的基础上，实现下载链接保护。当有数据下载需求时，展示变换后的文件存储地址，防止文件存储地址泄露，保护下载链接的安全。

[0084] 所述脱敏管理模块包括对数据脱敏算法进行添加、删除和修改的脱敏算法管理模块、对数据流通过程中的数据脱敏策略进行增加、删除和修改的脱敏策略管理模块、对敏感内容和敏感数据特征分析的基础上训练得到机器学习模型并在有敏感数据自动发现需求时对敏感数据进行自动发现的自学习引擎、对数据流通过程中的各个环节的数据脱敏进行实时监控以便及时发现异常的脱敏监控模块、对数据流通和交易过程中的数据脱敏任务进行分析审计的脱敏审计模块、对数据流通和交易过程中敏感数据保护处理进行评价的脱敏效能评估模块。

[0085] 优选地，所述脱敏算法管理模块向系统添加新的脱敏算法及相应的算法描述并可删除不适用的脱敏算法；所述脱敏策略管理模块对系统中的脱敏策略的使用情况进行挖掘分析添加用户常用的脱敏策略，修改不合适的脱敏策略并删除无用脱敏策略；所述脱敏监控模块对数据流通和交易中敏感数据的脱敏状态、脱敏策略、脱敏结果、数据需求方进行关联分析和挖掘，及时发现数据脱敏过程中的异常，以便在发生数据泄露是能实现数据追溯；所述脱敏效能评估模块对数据流通和交易中敏感数据脱敏保护的功能、敏感数据脱敏保护的应用场景、敏感数据脱敏保护的正确性和真实性以及有效性进行综合评价。

[0086] 具体地说，脱敏管理模块对数据脱敏算法与脱敏策略进行管理。包括脱敏算法管理与脱敏策略管理。数据脱敏算法管理模块具有对数据脱敏算法进行添加、删除、修改的功能。系统管理员可向系统中添加新的算法，对算法进行描述，并上传算法的jar包，可删除不常用的脱敏算法，还可以修改已有的数据脱敏算法的相关信息及jar包。脱敏策略管理模块具有对数据流通过程中的数据脱敏策略进行增加、删除、修改的功能。系统管理员可以通过对系统中策略的使用情况进行挖掘分析，添加用户常用的策略，修改不合适的策略，删除无

用的策略。自学习引擎是在对敏感内容、敏感数据特征分析的基础上，训练得到的机器学习模型，当有敏感数据自动发现需求时，调用自学习引擎进行敏感数据的自动发现。脱敏监控模块主要实现数据流通过程中各个环节的数据脱敏监控。系统管理员通过对脱敏任务的时间、进度、执行状态等信息进行实时监控，及时发现异常情况。脱敏审计模块是对数据流通过程中数据脱敏任务的分析，实现脱敏的审计。系统通过对数据脱敏日志中的数据、数据需求方、脱敏状态、脱敏策略、脱敏结果等信息进行关联分析与挖掘，及时发现数据脱敏过程中的异常，发生数据泄露事故时，能实现数据追溯，使责任到人。效能评估是对敏感数据保护方法的评价。该体系主要从敏感数据保护方法的功能、敏感数据保护的应用场景、敏感数据保护效果的正确性、真实性、有效性等多个方面对敏感数据保护方法进行综合评价，以支持敏感数据保护体系。

[0087] 如图2所示，一种大数据平台的数据流通和交易的敏感数据保护方法，采用大数据平台的数据流通和交易的敏感数据保护系统来实现，包括如下步骤：

[0088] 步骤一、资源发布人在数据采集前发现数据中的敏感内容，对数据采集中的敏感数据进行保护处理再将数据上传到大数据平台；

[0089] 步骤二、数据上传到大数据平台后，资源发布人对数据进行共享时，对数据共享中的敏感数据进行保护之后再共享发布；

[0090] 步骤三、数据需求方对资源发布人共享的数据请求交换，资源发布人对数据需求方请求交换的数据中的敏感内容进行定义，对数据交换中的敏感数据进行保护处理后再交换给数据需求方；

[0091] 步骤四、数据需求方请求下载资源发布人共享的数据，对数据下载中的敏感数据存储地址进行保护。

[0092] 如图3所示，所述对数据采集中的敏感数据进行保护包括如下步骤：

[0093] s11、资源发布人在向大数据平台上传数据前，制定元数据提取标准；进入步骤s12；

[0094] s12、解析上传的数据，进入步骤s13；

[0095] s13、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的元数据提取模块根据所述步骤s11中制定的元数据提取标准提取经步骤s12解析后的上传数据的元数据信息，进入步骤s14；

[0096] s14、根据所述步骤s13中提取的元数据信息进行敏感内容配置，进入步骤s15；

[0097] s15通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第一敏感数据发现模块根据所述步骤s14中配置的敏感内容，对数据中的敏感内容进行自动识别，进入步骤s16；

[0098] s16、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第一脱敏策略配置模块根据所述步骤s15中识别确定的敏感内容的特点，自动配置脱敏算法，形成脱敏策略，进入步骤s17；

[0099] s17、根据所述步骤s16中配置的脱敏算法，对数据进行脱敏处理，进入步骤s18；

[0100] s18、将脱敏后的数据上传到大数据平台。

[0101] 具体地说，数据采集首先要制定元数据提取的标准，即需要提取数据元素特征包括哪些方面；然后，解析待上传数据的类型（文档、传统数据库、实时数据库或者其他）；接着

提取出数据名称、数据摘要、资源拥有者、关键词、数据分类、数据标识、内容、背景、数据结构、存储位置等信息，为敏感数据的保护提供数据准备；依据提取的元数据信息，用户设定敏感内容、敏感内容所在的存储位置、敏感内容的摘要信息以及敏感数据的标识；依据敏感内容配置，采用基于规则的方式对数据中的敏感内容进行自动识别；用户也可跳过敏感内容设定，直接选择数据挖掘的方式进行敏感内容的发现，主要过程是提取数据的敏感内容特征，采用机器学习模型自动识别出敏感内容；在确定敏感内容的基础上，系统依据敏感内容的特点，自动推荐脱敏算法；如果推荐的脱敏算法不恰当，可以依据脱敏内容的脱敏类型，选择合适的脱敏算法；为敏感内容配置算法后，形成脱敏策略集；系统调用脱敏引擎，将预定义的敏感数据和脱敏策略集作为输入，执行脱敏；最后将生成的脱敏结果返回给用户。

[0102] 如图4所示，所述对数据共享中的敏感数据进行保护包括如下步骤：

[0103] s21、资源发布人上传数据到大数据平台，进入步骤s22；

[0104] s22、在大数据平台环境下，通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第二敏感数据发现模块根据数据属性定义敏感数据，进入步骤s23；

[0105] s23、根据所述步骤s22中定义的敏感数据确定数据中的敏感数据，进入步骤s24；

[0106] s24、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第二脱敏策略配置模块为所述步骤s23中确定的敏感数据配置脱敏算法，进入步骤s25；

[0107] s25、判断为敏感数据配置的脱敏算法是否合适，若否，进入步骤s26，若是，进入步骤s27；

[0108] s26、为敏感数据重新配置脱敏算法，进入步骤s27；

[0109] s27、根据配置的脱敏算法形成脱敏策略，进入步骤s28；

[0110] s28、保存脱敏策略，并对不同脱敏策略的使用情况进行统计分析，进入步骤s29；

[0111] s29、根据步骤s28中对脱敏策略使用情况的统计分析结果对脱敏策略使用率进行排序，进入步骤s210；

[0112] s210、根据脱敏策略使用率建立脱敏策略自动推荐机制。

[0113] 如图5-6所示，所述对数据交换中的敏感数据进行保护包括如下步骤：

[0114] s31、数据需求方输入查询条件对数据进行查询，进入步骤s32；

[0115] s32、对与查询条件匹配的数据进行差分隐私保护处理，进步步骤s33；

[0116] s33、将经过差分隐私保护处理的数据反馈给数据需求方，进步步骤s34；

[0117] s34、将反馈给数据需求方的数据以脱敏方式将数据显示给数据需求方供其查看，进入步骤s35；

[0118] s35、数据需求方请求共享数据，进入步骤s36；

[0119] s36、资源发布人审核数据需求方的身份和数据使用权限，以此为基础通过所述大数据平台的数据流通与交易的敏感数据保护系统中的敏感数据定义模块预定义相对于数据需求方的敏感数据，进入步骤s37；

[0120] s37、判断预定义的敏感数据与数据需求方身份与权限是否相符，若否，进入步骤s38，若是，进入步骤s39；

[0121] s38、重新定义敏感数据，进入步骤s39；

[0122] s39、根据定义确定相对于数据需求方的敏感数据，进入步骤s310；

[0123] s310、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的第三脱敏

策略配置模块判断确定的敏感数据的脱敏策略是否合适，若否，进入步骤s311，若是，进入步骤s312；

[0124] s311、修改脱敏策略，进入步骤s312；

[0125] s312、确认敏感数据脱敏策略，进入步骤s313；

[0126] s313、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的数据脱敏引擎根据确认的脱敏策略对数据需求方请求共享的数据进行脱敏处理，进入步骤s314；

[0127] s314、通过所述大数据平台的数据流通与交易的敏感数据保护系统中的脱敏验证模块校验脱敏数据的正确性，进入步骤s315；

[0128] s315、向数据需求方展示校验后的可供共享的脱敏数据。

[0129] 如图7所示，所述对数据下载中的敏感数据存储地址进行保护包括如下步骤：

[0130] S41、数据需求方从展示的结果中选择需要下载的数据文件，进入步骤s42；

[0131] S42、对数据需求方选择下载的文件进行下载链接保护处理，生成新的可映射到文件原存储地址的安全链接，进入步骤s43；

[0132] s43、向数据需求方展示新生成的安全链接供其下载文件。

[0133] 优选的，通过所述大数据平台的数据流通与交易的敏感数据保护系统的脱敏管理模块对每一环节的脱敏算法、脱敏策略进行管理；

[0134] 通过所述脱敏管理模块中的脱敏算法管理模块以实现为敏感数据配置合适脱敏算法的目的对每一环节的脱敏算法配置进行添加、删除和修改处理；

[0135] 通过所述脱敏管理模块中的脱敏策略管理模块以实现为敏感数据配置合适脱敏策略的目的对数据流通和交易中的脱敏策略进行增加、删除和修改管理；

[0136] 通过所述脱敏管理模块中的自学习引擎对数据流通和交易中的敏感内容和敏感数据特征进行分析训练以得到机器学习模型满足敏感数据的自动发现需求；

[0137] 通过所述脱敏管理模块中的脱敏监控模块对数据流通中的各个环节的数据脱敏进行实时监控以便及时发现异常情况；

[0138] 通过所述脱敏管理模块中的脱敏审计模块对数据流通和交易中的数据脱敏任务进行分析审计；

[0139] 通过所述脱敏管理模块中的脱敏效能评估模块对数据流通和交易中的敏感数据的保护处理进行评价。

[0140] 当资源发布人将数据发布到大数据平台后，在进入资源共享前，需要进行敏感数据发现与脱敏策略配置，具体地说，数据上传到系统后，资源发布人对数据进行共享时，首先判断数据的属性的大小，如果数据的属性比较少，依据个人的经验和相关规定定义敏感数据；当数据的属性较多，采用基于专家系统或者自然语言处理的方式。基于专家系统的方式是依据敏感数据与文本处理领域专家的知识与经验，分析并总结敏感数据的规律，形成敏感数据发现的规则，并利用这些规则对数据是否敏感进行推理与判断，自动发现敏感数据。基于自然语言处理的敏感数据自动发现是对平台的不同领域敏感数据进行深度分析，采用特征提取算法提取敏感数据的特征，在敏感数据特征提取的基础上，采用自然语言处理模型，实现对身份证号、银行卡号、地址、出生日期、公司名称、金额、口令、姓名等敏感数据的自动发现。定义好敏感数据后要确定数据中的敏感数据。系统为已确定的敏感数据进行脱敏算法的自动推荐；如果推荐的脱敏算法不合适，可以进行修改，修改后形成新的脱敏

策略；当系统中策略达到一定数量后，对策略的使用情况进行统计分析，得到用户使用较多的策略，形成策略排序；系统依据策略排序，为用户自动推荐策略。

[0141] 资源发布人将数据发布后，数据需求方输入查询条件对数据进行查询检索；系统依据查询条件和数据本身的特点，对原数据原始数据、原始数据的转换、或者是对统计结果使用拉普拉斯机制和指数机制的方法实现差分隐私，如果为数值型数据通过拉普拉斯机制对结果添加噪声实现差分隐私保护，如果为非数值型数据通过指数机制对结果添加噪声实现差分隐私保护；查询后，系统将结果反馈给数据需求方；当数据需求方查看数据的相关信息时，对数据信息中的数据来源、数据摘要、数据所有者等敏感信息，采用遮挡、置换等方式进行脱敏。

[0142] 数据需求方在查询结果中申请需要共享的数据，资源发布人依据数据需求方的身份和数据使用权限，定义敏感数据；如果预定义的敏感数据与数据需求方身份与权限不符，修改敏感数据；然后确定数据需求方请求共享的数据中的敏感数据；系统在敏感数据定义模块的基础上，结合敏感数据的特征，对数据共享过程中预定义的脱敏策略进行确认；如果敏感策略不合适进行修改，如果合适进行确认，形成最终的脱敏策略；当数据需求方发出数据交换请求时，平台调用预定义的敏感数据及脱敏策略，使用数据脱敏引擎对数据进行脱敏处理；系统将脱敏前数据与脱敏后数据进行数据格式、长度、完整性的检查，并使用相应脱敏规则对数据进行脱敏，将结果与脱敏后的数据进行比对，验证数据脱敏的正确性；通过数据视图展示原数据之间的逻辑关系；依原数据之间的逻辑关系查询脱敏后数据的数据项是否仍存在逻辑性；依据实际数据脱敏需求，为脱敏前后的数据分布统计设置阈值；对比脱敏前后数据分布的统计的平均数、标准差、中位数、统计分布图，如果平均数、标准差、中位数超过设定阈值，统计形状差异过大，需重新调整脱敏策略。

[0143] 本发明的有益效果在于：(1)从数据流通的整个环节实现了敏感数据的保护；(2)在数据流通的不同环节使用不同的敏感数据保护方法，实现敏感数据保护与数据可用性之间的平衡；(3)提出了基于专家系统和自然语言处理的敏感数据自动发现方法，能够自动发现敏感数据；(4)提出了验证脱敏结果正确性与真实性的方法，能够有效地度量数据脱敏环节。

[0144] 以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明披露的技术范围内，根据本发明的技术方案及其发明构思加以等同替换或改变，都应涵盖在本发明的保护范围之内。

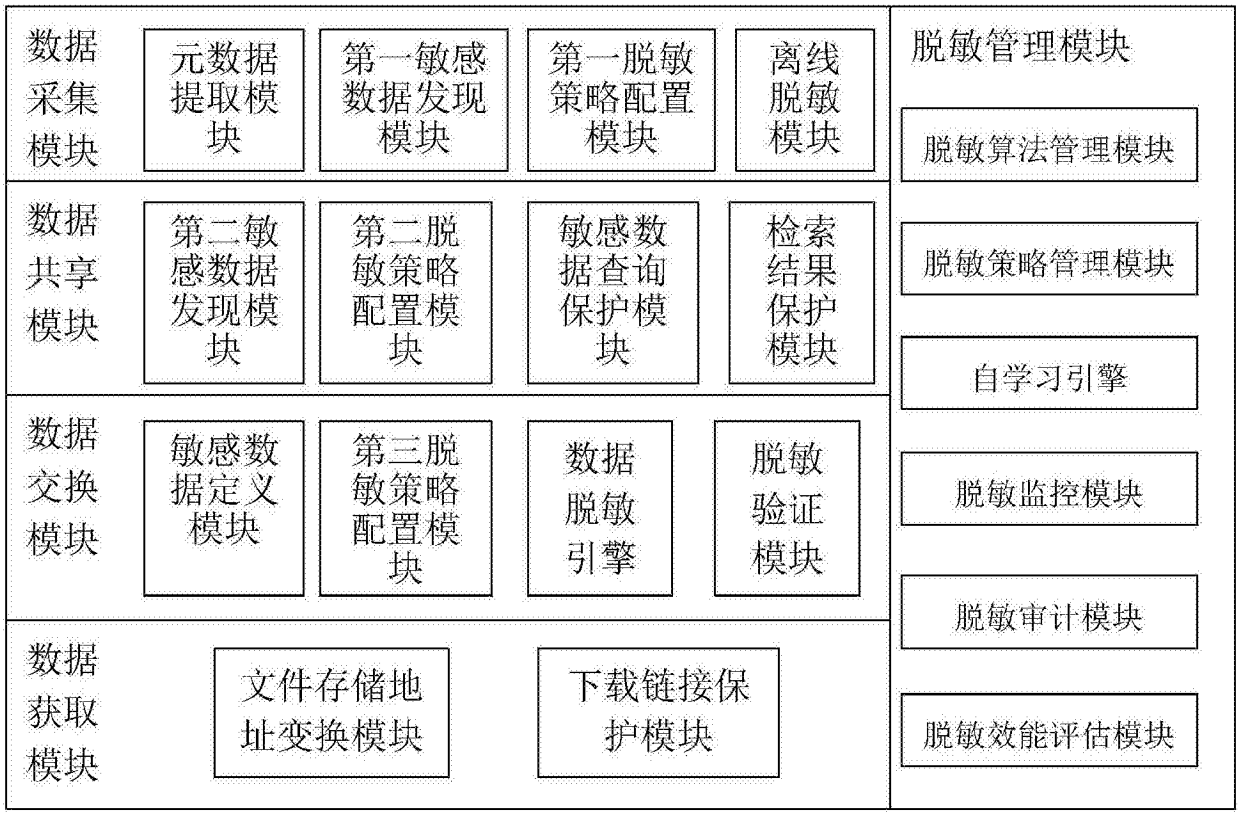

数据采集模块
元数据提取模块
第一敏感数据发现模块
第一脱敏策略配置模块
离线脱敏模块
数据共享模块
第二敏感数据发现模块
第二脱敏策略配置模块
敏感数据查询保护模块
检索结果保护模块
数据交换模块
敏感数据定义模块
第三脱敏策略配置模块
数据脱敏引擎
脱敏验证模块
数据获取模块
文件存储地址变换模块
下载链接保护模块
脱敏管理模块
脱敏算法管理模块
脱敏策略管理模块
自学习引擎
脱敏监控模块
脱敏审计模块
脱敏效能评估模块

图1

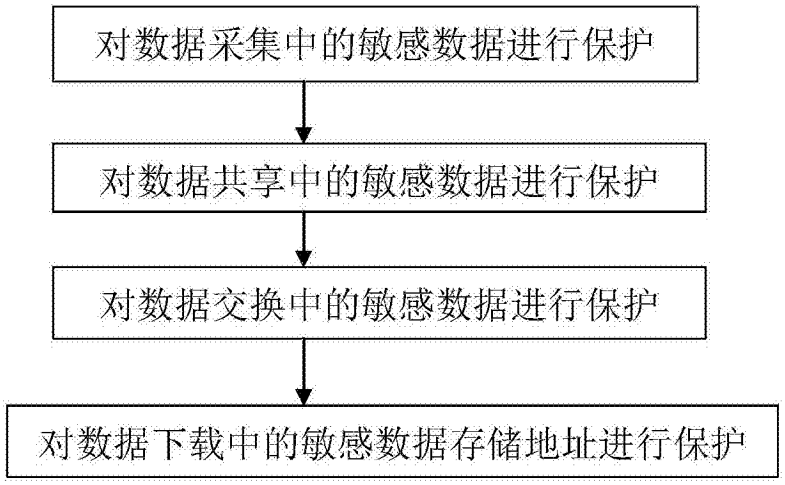

对数据采集中的敏感数据进行保护
对数据共享中的敏感数据进行保护
对数据交换中的敏感数据进行保护
对数据下载中的敏感数据存储地址进行保护

图2

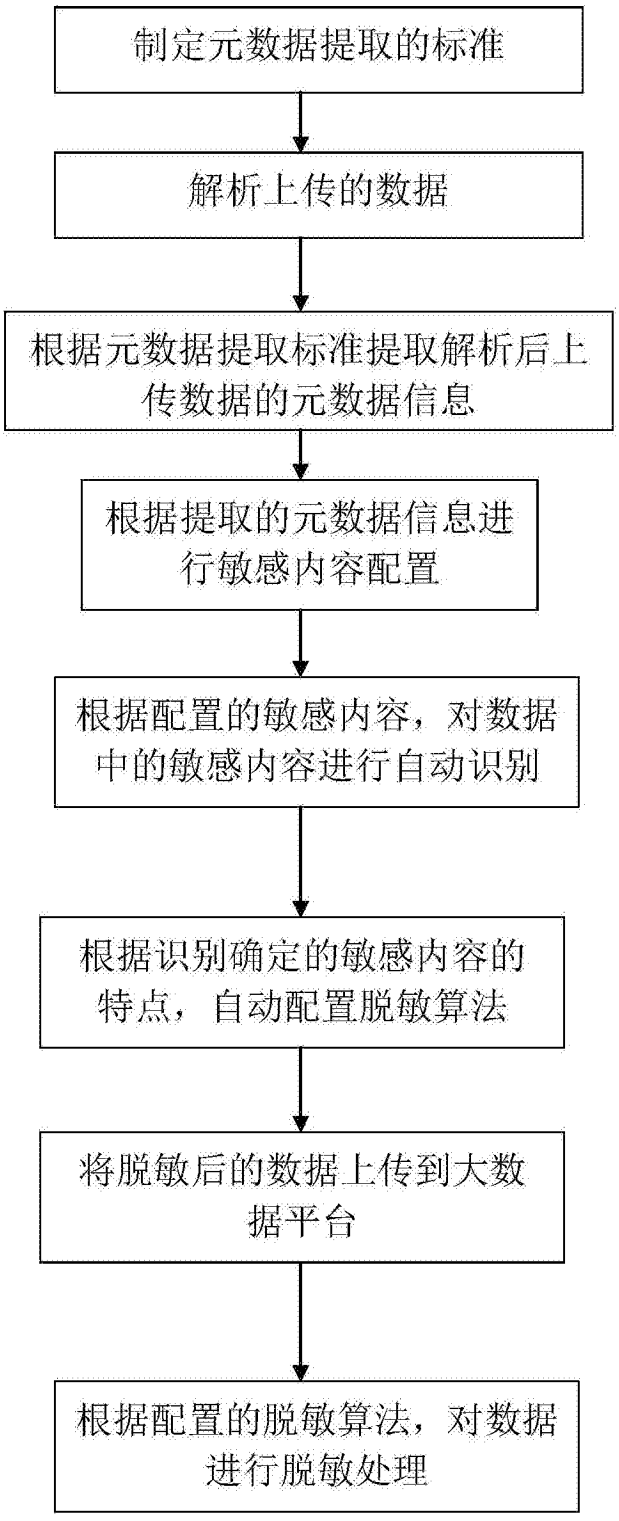

制定元数据提取的标准
解析上传的数据
根据元数据提取标准提取解析后上传数据的元数据信息
根据提取的元数据信息进行敏感内容配置
根据配置的敏感内容，对数据中的敏感内容进行自动识别
根据识别确定的敏感内容的特点，自动配置脱敏算法
将脱敏后的数据上传到大数据平台
根据配置的脱敏算法，对数据进行脱敏处理

图3

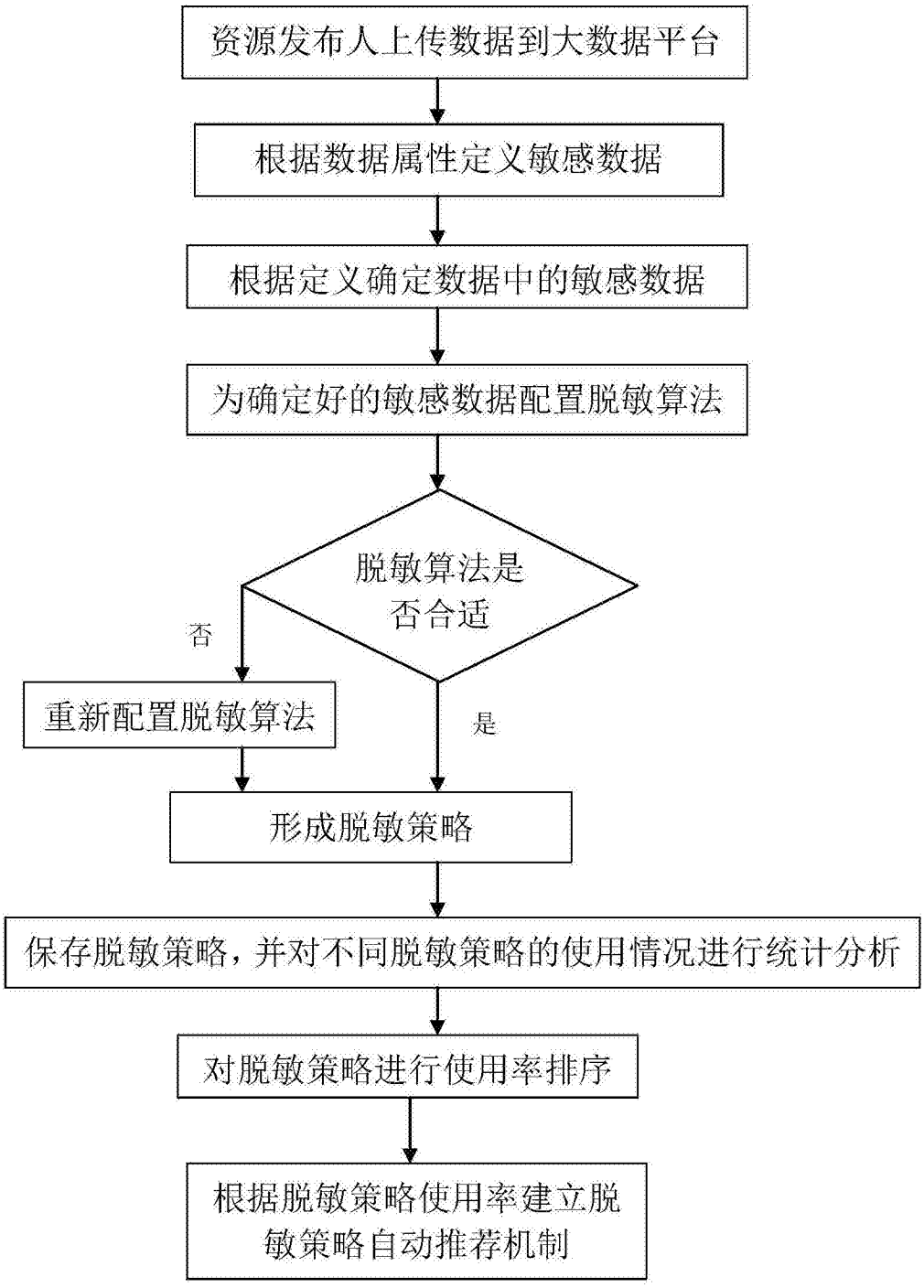

资源发布人上传数据到大数据平台
根据数据属性定义敏感数据
根据定义确定数据中的敏感数据
为确定好的敏感数据配置脱敏算法
脱敏算法是否合适
否
是
重新配置脱敏算法
形成脱敏策略
保存脱敏策略，并对不同脱敏策略的使用情况进行统计分析
对脱敏策略进行使用率排序
根据脱敏策略使用率建立脱敏策略自动推荐机制

图4

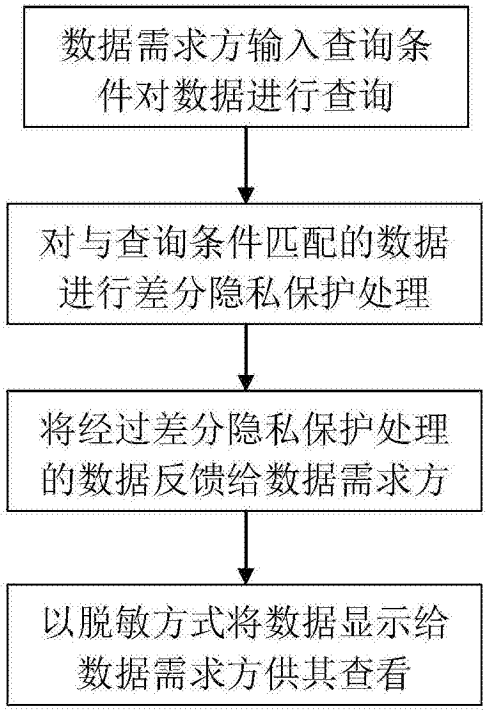

数据需求方输入查询条件对数据进行查询
对与查询条件匹配的数据进行差分隐私保护处理
将经过差分隐私保护处理的数据反馈给数据需求方
以脱敏方式将数据显示给数据需求方供其查看

图5

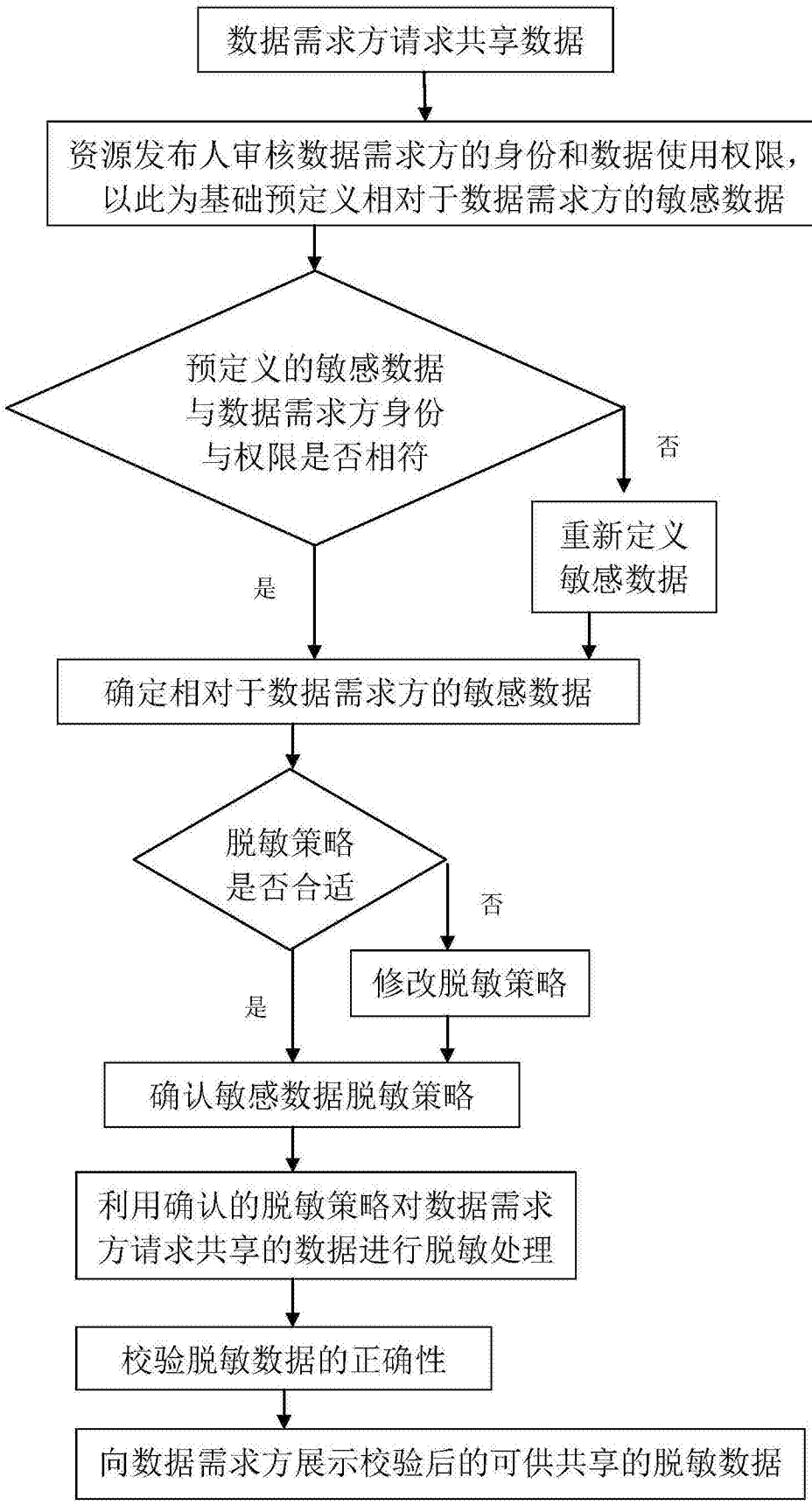

数据需求方请求共享数据
资源发布人审核数据需求方的身份和数据使用权限，以此为基础预定义相对于数据需求方的敏感数据
预定义的敏感数据与数据需求方身份与权限是否相符
否
是
重新定义敏感数据
确定相对于数据需求方的敏感数据
脱敏策略是否合适
否
是
修改脱敏策略
确认敏感数据脱敏策略
利用确认的脱敏策略对数据需求方请求共享的数据进行脱敏处理
校验脱敏数据的正确性
向数据需求方展示校验后的可供共享的脱敏数据

图6

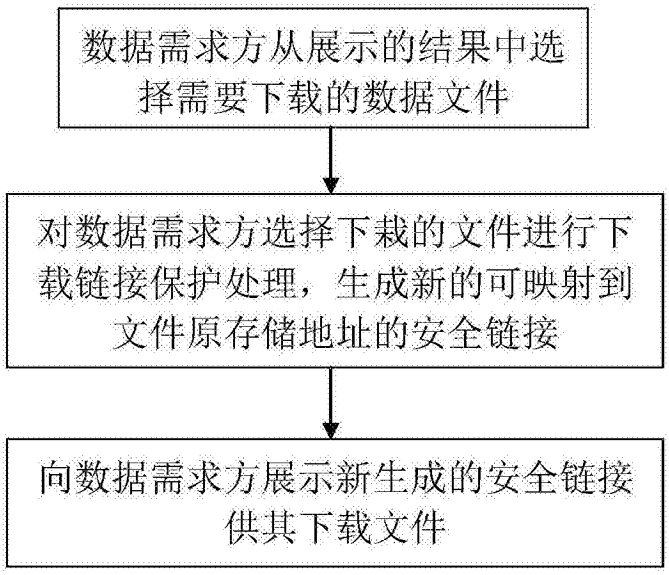

数据需求方从展示的结果中选择需要下载的数据文件
对数据需求方选择下载的文件进行下载链接保护处理，生成新的可映射到文件原存储地址的安全链接
向数据需求方展示新生成的安全链接供其下载文件

图7