



(12)发明专利

(10)授权公告号 CN 104217136 B

(45)授权公告日 2017.05.03

(21)申请号 201310220475.7

(22)申请日 2013.06.05

(65)同一申请的已公布的文献号  
申请公布号 CN 104217136 A

(43)申请公布日 2014.12.17

(73)专利权人 北京齐尔布莱特科技有限公司  
地址 100080 北京市海淀区丹棱街3号B座  
10层1010室

(72)发明人 韩华

(74)专利代理机构 北京市盛峰律师事务所  
11337

代理人 赵建刚

(51)Int.Cl.  
G06F 21/10(2013.01)  
G06F 17/30(2006.01)

(56)对比文件

CN 102916935 A,2013.02.06,  
CN 101267299 A,2008.09.17,  
CN 1505309 A,2004.06.16,  
CN 101814118 A,2010.08.25,  
CN 102156953 A,2011.08.17,

审查员 简文雨

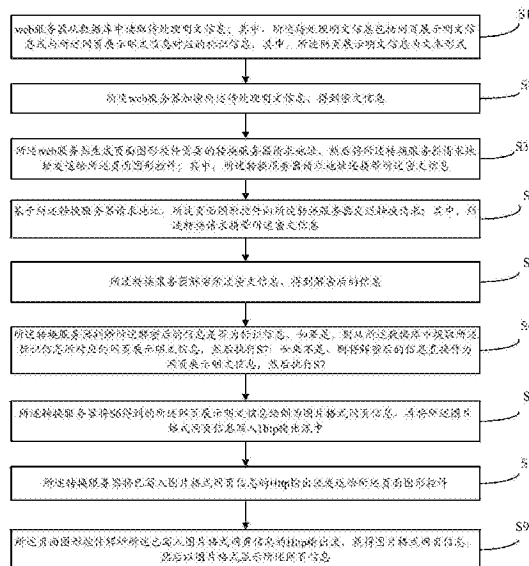
权利要求书2页 说明书4页 附图2页

(54)发明名称

防止网页文本信息被自动抓取的方法和系统

(57)摘要

本发明提供一种防止网页文本信息被自动抓取的方法和系统,方法包括:页面图形控件向转换服务器发送转换请求;转换服务器解密密文信息,得到解密后的信息,判断解密后的信息是否为标识信息,如果是,则从数据库中提取标识信息所对应的网页展示明文信息,如果不是,则将解密后的信息直接作为网页展示明文信息;将得到的网页展示明文信息绘制为图片格式网页信息,再写入Http输出流中;将Http输出流发送给页面图形控件;页面图形控件解析Http输出流,获得图片格式网页信息,然后以图片格式显示所述网页信息。以图片格式展示网站所发布的信息,从而达到防止网站所发布的信息被自动抓取的效果。



1. 一种防止网页文本信息被自动抓取的方法,其特征在于,包括以下步骤:

S1,web服务器从数据库中读取待处理明文信息;其中,所述待处理明文信息包括网页展示明文信息或与所述网页展示明文信息对应的标识信息;其中,所述网页展示明文信息为文本形式;

S2,所述web服务器加密所述待处理明文信息,得到密文信息;

S3,所述web服务器生成页面图形控件需要的转换服务器请求地址,然后将所述转换服务器请求地址发送给所述页面图形控件;其中,所述转换服务器请求地址还携带所述密文信息;

S4,基于所述转换服务器请求地址,所述页面图形控件向所述转换服务器发送转换请求;其中,所述转换请求携带所述密文信息;

S5,所述转换服务器解密所述密文信息,得到解密后的信息;

S6,所述转换服务器判断所述解密后的信息是否为标识信息,如果是,则从所述数据库中提取所述标识信息所对应的网页展示明文信息,然后执行S7;如果不是,则将解密后的信息直接作为网页展示明文信息,然后执行S7;

S7,所述转换服务器将S6得到的所述网页展示明文信息绘制为图片格式网页信息,再将所述图片格式网页信息写入Http输出流中;

S8,所述转换服务器将已写入图片格式网页信息的Http输出流发送给所述页面图形控件;

S9,所述页面图形控件解析所述已写入图片格式网页信息的Http输出流,获得图片格式网页信息,然后以图片格式显示所述网页信息;

其中,S2具体为:

所述web服务器使用DES算法加密所述待处理明文信息,得到密文信息;

其中,S9中,所述页面图形控件以图片格式显示所述网页信息具体为:

所述页面图形控件预先定义需要显示的图片的属性;然后基于所述图片的属性,以图片格式显示所述网页信息;

其中,所述图片的属性包括图片的长度信息、图片的宽度信息、图片的分辨率信息、图片中字符显示的位置信息、图片中字符颜色信息、图片中字符字体信息和图片中字符大小信息中的一种或几种。

2. 一种防止网页文本信息被自动抓取的系统,其特征在于,包括:web服务器、页面图形控件和转换服务器;

所述web服务器用于:从数据库中读取待处理明文信息;其中,所述待处理明文信息包括网页展示明文信息或与所述网页展示明文信息对应的标识信息;其中,所述网页展示明文信息为文本形式;加密所述待处理明文信息,得到密文信息;生成页面图形控件需要的转换服务器请求地址,然后将所述转换服务器请求地址发送给所述页面图形控件;其中,所述转换服务器请求地址还携带所述密文信息;

所述页面图形控件用于:接收所述web服务器发送的转换服务器请求地址,基于所述转换服务器请求地址,向所述转换服务器发送转换请求;其中,所述转换请求携带所述密文信息;

所述转换服务器用于:接收所述页面图形控件发送的转换请求,解密所述密文信息,得

到解密后的信息;判断所述解密后的信息是否为标识信息,如果是,则从所述数据库中提取所述标识信息所对应的网页展示明文信息,如果不是,则将解密后的信息直接作为网页展示明文信息;然后,将得到的网页展示明文信息绘制为图片格式网页信息,再将所述图片格式网页信息写入Http输出流中,将已写入图片格式网页信息的Http输出流发送给所述页面图形控件;

所述页面图形控件还用于:接收所述转换服务器发送的Http输出流,解析所述已写入图片格式网页信息的Http输出流,获得图片格式网页信息,然后以图片格式显示所述网页信息。

## 防止网页文本信息被自动抓取的方法和系统

### 技术领域

[0001] 本发明属于信息处理技术领域,具体涉及一种防止网页文本信息被自动抓取的方法和系统。

### 背景技术

[0002] 在当前网络信息时代,用户常常向特定网站的网页发布文本格式信息,例如:证件号码、手机号、商品价格等。现有技术中,出现了网络蜘蛛或类似的网页信息自动抓取工具,通过该种工具,能够自动获取网页中展示的文本格式信息。

[0003] 但是,以商品价格为例,当网站A发布商品A的价格后,其他网站通过自动抓取工具进行全网抓取,从而抓取到众多网站所发布的商品A价格,然后对商品A价格进行比对,制定商品A经营策略,从而不利于网站A。

[0004] 因此,如何有效保护用户在网站所发布的信息,防止其被网页信息自动抓取工具所抓取到,具有重要意义。

### 发明内容

[0005] 针对现有技术存在的缺陷,本发明提供一种防止网页文本信息被自动抓取的方法和系统,以图片格式展示网站所发布的信息,从而达到防止网站所发布的信息被自动抓取的效果。

[0006] 本发明采用的技术方案如下:

[0007] 本发明提供一种防止网页文本信息被自动抓取的方法,包括以下步骤:

[0008] S1,web服务器从数据库中读取待处理明文信息;其中,所述待处理明文信息包括网页展示明文信息或与所述网页展示明文信息对应的标识信息;其中,所述网页展示明文信息为文本形式;

[0009] S2,所述web服务器加密所述待处理明文信息,得到密文信息;

[0010] S3,所述web服务器生成页面图形控件需要的转换服务器请求地址,然后将所述转换服务器请求地址发送给所述页面图形控件;其中,所述转换服务器请求地址还携带所述密文信息;

[0011] S4,基于所述转换服务器请求地址,所述页面图形控件向所述转换服务器发送转换请求;其中,所述转换请求携带所述密文信息;

[0012] S5,所述转换服务器解密所述密文信息,得到解密后的信息;

[0013] S6,所述转换服务器判断所述解密后的信息是否为标识信息,如果是,则从所述数据库中提取所述标识信息所对应的网页展示明文信息,然后执行S7;如果不是,则将解密后的信息直接作为网页展示明文信息,然后执行S7;

[0014] S7,所述转换服务器将S6得到的所述网页展示明文信息绘制为图片格式网页信息,再将所述图片格式网页信息写入Http输出流中;

[0015] S8,所述转换服务器将已写入图片格式网页信息的Http输出流发送给所述页面图

形控件；

[0016] S9,所述页面图形控件解析所述已写入图片格式网页信息的Http输出流,获得图片格式网页信息,然后以图片格式显示所述网页信息。

[0017] 优选的,S2具体为:

[0018] 所述web服务器使用DES算法加密所述待处理明文信息,得到密文信息。

[0019] 优选的,S9中,所述页面图形控件以图片格式显示所述网页信息具体为:

[0020] 所述页面图形控件预先定义需要显示的图片的属性;然后基于所述图片的属性,以图片格式显示所述网页信息。

[0021] 优选的,所述图片的属性包括图片的长度信息、图片的宽度信息、图片的分辨率信息、图片中字符显示的位置信息、图片中字符颜色信息、图片中字符字体信息和图片中字符大小信息中的一种或几种。

[0022] 本发明还提供一种防止网页文本信息被自动抓取的系统,包括:web服务器、页面图形控件和转换服务器;

[0023] 所述web服务器用于:从数据库中读取待处理明文信息;其中,所述待处理明文信息包括网页展示明文信息或与所述网页展示明文信息对应的标识信息;其中,所述网页展示明文信息为文本形式;加密所述待处理明文信息,得到密文信息;生成页面图形控件需要的转换服务器请求地址,然后将所述转换服务器请求地址发送给所述页面图形控件;其中,所述转换服务器请求地址还携带所述密文信息;

[0024] 所述页面图形控件用于:接收所述web服务器发送的转换服务器请求地址,基于所述转换服务器请求地址,向所述转换服务器发送转换请求;其中,所述转换请求携带所述密文信息;

[0025] 所述转换服务器用于:接收所述页面图形控件发送的转换请求,解密所述密文信息,得到解密后的信息;判断所述解密后的信息是否为标识信息,如果是,则从所述数据库中提取所述标识信息所对应的网页展示明文信息,如果不是,则将解密后的信息直接作为网页展示明文信息;然后,将得到的网页展示明文信息绘制为图片格式网页信息,再将所述图片格式网页信息写入Http输出流中,将已写入图片格式网页信息的Http输出流发送给所述页面图形控件;

[0026] 所述页面图形控件还用于:接收所述转换服务器发送的Http输出流,解析所述已写入图片格式网页信息的Http输出流,获得图片格式网页信息,然后以图片格式显示所述网页信息。

[0027] 本发明的有益效果如下:

[0028] 本发明提供一种防止网页文本信息被自动抓取的方法和系统,以图片格式展示网站所发布的信息,既能达到防止网站所发布的信息被自动抓取的效果;还具有安全、高效、简洁、易用的优点。

## 附图说明

[0029] 图1为本发明提供的防止网页文本信息被自动抓取的方法流程示意图;

[0030] 图2为本发明提供的防止网页文本信息被自动抓取的系统结构示意图。

## 具体实施方式

[0031] 以下结合附图对本发明进行详细说明：

[0032] 如图1所示，本发明提供一种防止网页文本信息被自动抓取的方法，包括以下步骤：

[0033] S1，web服务器从数据库中读取待处理明文信息；其中，所述待处理明文信息包括网页展示明文信息或与所述网页展示明文信息对应的标识信息；其中，所述网页展示明文信息为文本形式；

[0034] S2，所述web服务器加密所述待处理明文信息，得到密文信息；

[0035] 本发明对具体的加密算法并不限制，可以为DES算法。通过对待处理明文信息进行加密，可以提高web服务器到转换服务器之间信息传输的安全性。

[0036] S3，所述web服务器生成页面图形控件需要的转换服务器请求地址，然后将所述转换服务器请求地址发送给所述页面图形控件；其中，所述转换服务器请求地址还携带所述密文信息；

[0037] S4，基于所述转换服务器请求地址，所述页面图形控件向所述转换服务器发送转换请求；其中，所述转换请求携带所述密文信息；

[0038] S5，所述转换服务器解密所述密文信息，得到解密后的信息；

[0039] S6，所述转换服务器判断所述解密后的信息是否为标识信息，如果是，则从所述数据库中提取所述标识信息所对应的网页展示明文信息，然后执行S7；如果不是，则将解密后的信息直接作为网页展示明文信息，然后执行S7；

[0040] S7，所述转换服务器将S6得到的所述网页展示明文信息绘制为图片格式网页信息，再将所述图片格式网页信息写入Http输出流中；

[0041] 需要强调的是，本步骤中，将图片格式网页信息以流的形式写入Http输出流中，因此，转换服务器向页面图形控件返回的为数据流，而不是图片格式，由于流小于图片的大小，因此，可以提高数据传输的速度；另外，以流的方式传输，不容易被黑客破解，从而提高数据传输的安全性。

[0042] S8，所述转换服务器将已写入图片格式网页信息的Http输出流发送给所述页面图形控件；

[0043] S9，所述页面图形控件解析所述已写入图片格式网页信息的Http输出流，获得图片格式网页信息，然后以图片格式显示所述网页信息。

[0044] 具体的，页面图形控件预先定义需要显示的图片的属性；然后基于所述图片的属性，以图片格式显示所述网页信息。其中，图片的属性包括图片的长度信息、图片的宽度信息、图片的分辨率信息、图片中字符显示的位置信息、图片中字符颜色信息、图片中字符字体信息和图片中字符大小信息中的一种或几种。

[0045] 如图2所示，本发明还提供一种防止网页文本信息被自动抓取的系统，包括：web服务器、页面图形控件和转换服务器；

[0046] 所述web服务器用于：从数据库中读取待处理明文信息；其中，所述待处理明文信息包括网页展示明文信息或与所述网页展示明文信息对应的标识信息；其中，所述网页展示明文信息为文本形式；加密所述待处理明文信息，得到密文信息；生成页面图形控件需要

的转换服务器请求地址,然后将所述转换服务器请求地址发送给所述页面图形控件;其中,所述转换服务器请求地址还携带所述密文信息;

[0047] 所述页面图形控件用于:接收所述web服务器发送的转换服务器请求地址,基于所述转换服务器请求地址,向所述转换服务器发送转换请求;其中,所述转换请求携带所述密文信息;

[0048] 所述转换服务器用于:接收所述页面图形控件发送的转换请求,解密所述密文信息,得到解密后的信息;判断所述解密后的信息是否为标识信息,如果是,则从所述数据库中提取所述标识信息所对应的网页展示明文信息,如果不是,则将解密后的信息直接作为网页展示明文信息;然后,将得到的网页展示明文信息绘制为图片格式网页信息,再将所述图片格式网页信息写入Http输出流中,将已写入图片格式网页信息的Http输出流发送给所述页面图形控件;

[0049] 所述页面图形控件还用于:接收所述转换服务器发送的Http输出流,解析所述已写入图片格式网页信息的Http输出流,获得图片格式网页信息,然后以图片格式显示所述网页信息。

[0050] 综上所述,本发明提供的防止网页文本信息被自动抓取的方法和系统,以图片格式展示网站所发布的信息,既能达到防止网站所发布的信息被自动抓取的效果;还具有安全、高效、简洁、易用的优点。

[0051] 以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

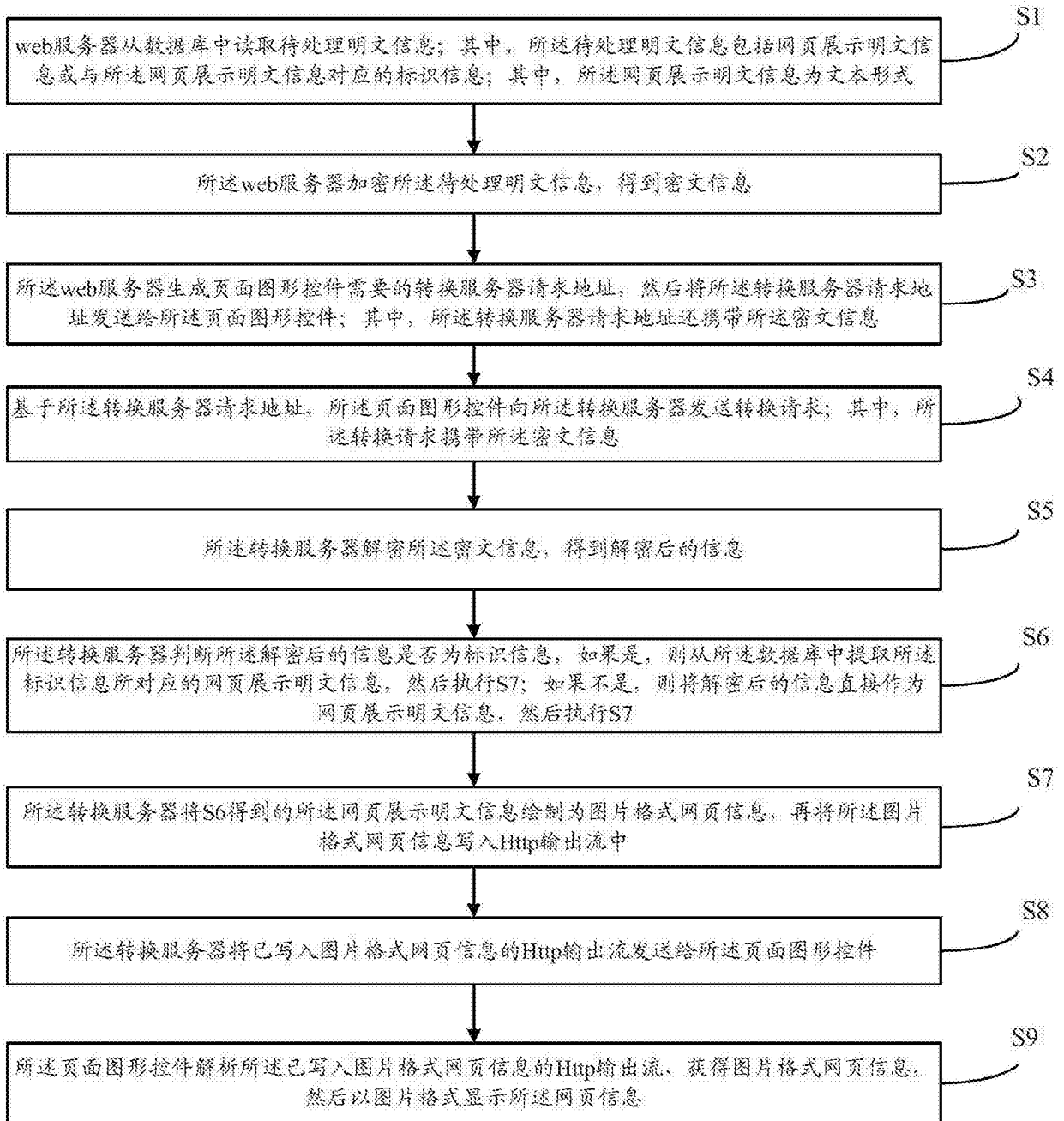


图1



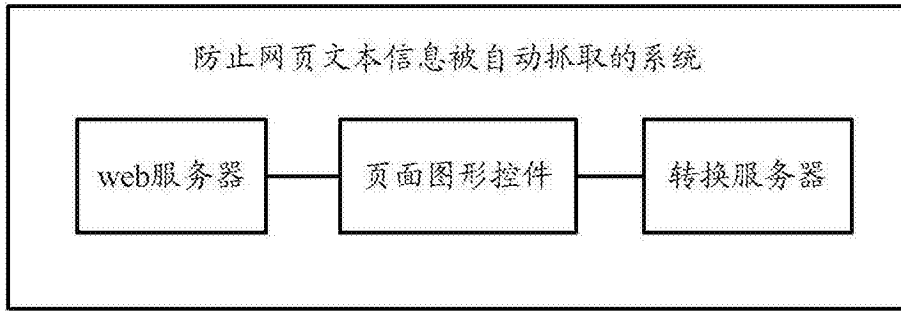


图2