



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G06F 16/9035 (2019.05); G06F 17/27 (2019.05); H04L 12/22 (2019.05)

(21)(22) Заявка: 2018105780, 12.07.2018

(24) Дата начала отсчета срока действия патента:
12.07.2018

Дата регистрации:
02.10.2019

Приоритет(ы):

(22) Дата подачи заявки: 12.07.2018

(45) Опубликовано: 02.10.2019 Бюл. № 28

Адрес для переписки:

143026, Москва, ул. территория
инновационного центра "Сколково", 4, ООО
"ЦИС "Сколково"

(72) Автор(ы):

Ашманов Игорь Станиславович (RU),
Иванов Алексей Петрович (RU),
Отарбиев Эльдар Отарбиевич (RU),
Пашко Дмитрий Алексеевич (RU),
Тихонов Максим Викторович (RU)

(73) Патентообладатель(и):

Акционерное Общество "Ремпаро" (RU)

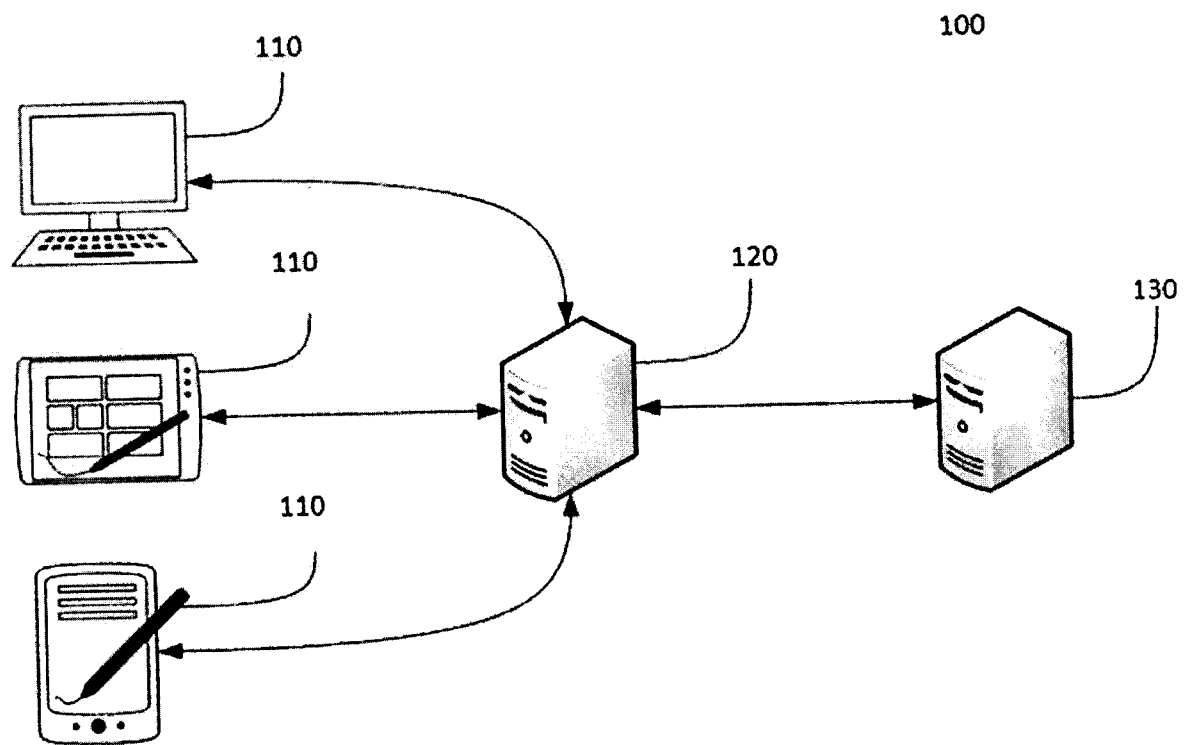
(56) Список документов, цитированных в отчете
о поиске: RU 2305314 C1, 27.08.2007. RU
2409849 C2, 20.01.2011. RU 2419858 C2,
27.05.2011. RU 2446460 C1, 27.03.2012. RU
2336561 C2, 20.10.2008. WO 2001/098934 A2,
27.12.2001. US 8601084 B2, 03.12.2013. US 2007/
0271220 A1, 22.11.2007.

(54) Способ использования системы определения тематики документов для целей информационной безопасности

(57) Реферат:

Изобретение относится к способу автоматического управления доступом к веб-странице. Технический результат заключается в обеспечении автоматического управления доступом к веб-странице. В способе формируют запрос на доступ к веб-странице в сети с устройств пользователей; внешний сервер получает запрос пользователя, формирует ответ и отправляет его обратно к пользователю; фильтрующий сервер перехватывает сформированный ответ для пользователя, осуществляет извлечение текстовой части из полученной веб-страницы, анализирует извлеченный текст запрошенного документа и разбивает текст на уровень предложений и

уровень слов; осуществляет лемматизацию полученных предложений и слов; сравнивает приведенные предложения и слова к словарной форме с базой классифицированных нежелательных слов, словосочетаний и цитат; формирует список тематик, к которым относится этот документ, который сравнивается с профилем доступа пользователя, в котором зафиксированы критерии, определяющие, может ли пользователь получить доступ к документу, отнесенному к данному набору тематик, или нет; по результатам сравнения блокируют или разрешают доступ к документу на веб-странице. 3 з.п. ф-лы, 4 ил.



Фиг. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC

G06F 16/9035 (2019.05); *G06F 17/27* (2019.05); *H04L 12/22* (2019.05)(21)(22) Application: **2018105780, 12.07.2018**(24) Effective date for property rights:
12.07.2018Registration date:
02.10.2019

Priority:

(22) Date of filing: **12.07.2018**(45) Date of publication: **02.10.2019** Bull. № 28

Mail address:

**143026, Moskva, ul. territoriya innovatsionnogo
tsentra "Skolkovo", 4, OOO "TSIS "Skolkovo"**

(72) Inventor(s):

**Ashmanov Igor Stanislavovich (RU),
Ivanov Aleksey Petrovich (RU),
Otarbiev Eldar Otarbievich (RU),
Pashko Dmitrij Alekseevich (RU),
Tikhonov Maksim Viktorovich (RU)**

(73) Proprietor(s):

Aksionernoe Obshchestvo "Remparo" (RU)(54) **METHOD OF USING DOCUMENT IDENTIFICATION SYSTEM FOR INFORMATION SECURITY PURPOSES**

(57) Abstract:

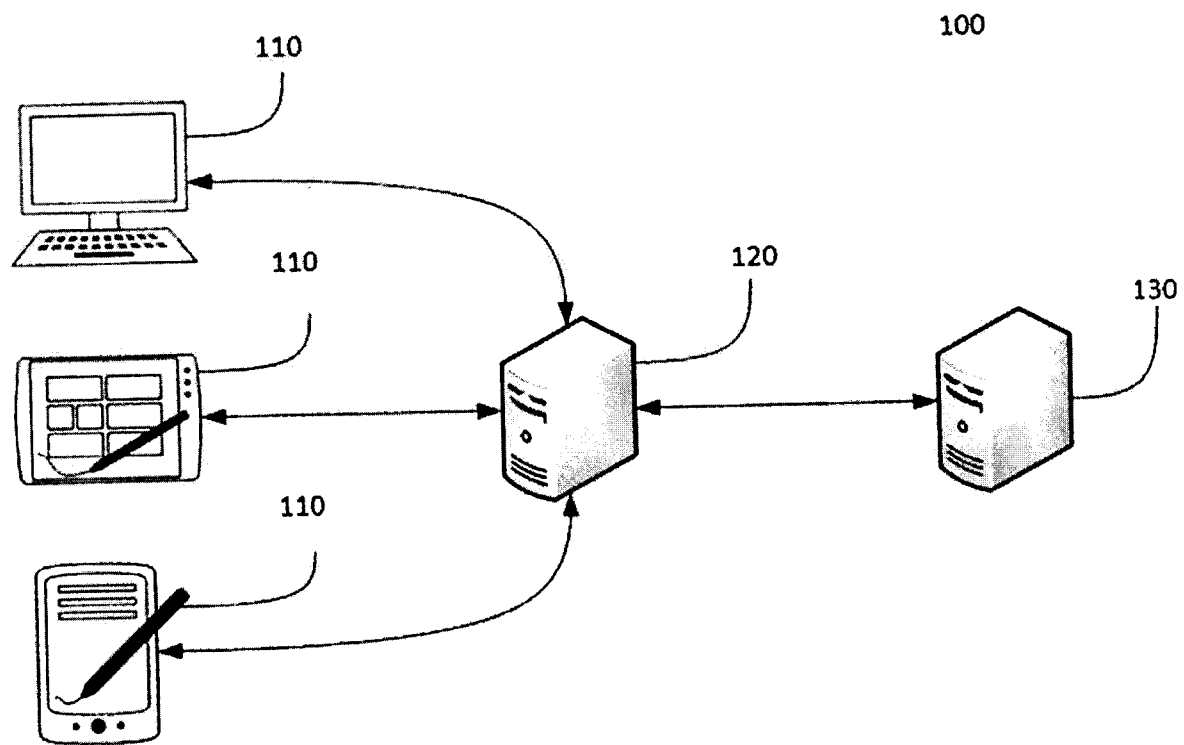
FIELD: physics.

SUBSTANCE: invention relates to a method of automatically controlling access to a web page. Method comprises generating a request to access a web page in a network from user devices; external server receives user request, generates response and sends it back to user; filtering server intercepts generated response for user, extracts text part from received web page, analyses extracted text of requested document and splits text into sentence level and word level; performs lemmatization of obtained sentences and words; compares presented

sentences and words to dictionary form with base of classified undesirable words, collocations and citations; forms a list of topics, to which this document relates, which is compared with a user access profile, in which criteria are fixed, determining whether a user can access a document related to a given set of topics, or not; results of comparison are used to block or allow access to a document on a web page.

EFFECT: technical result is providing automatic control of access to a web page.

4 cl, 4 dwg



Фиг. 1

ОБЛАСТЬ ТЕХНИКИ

Настоящее техническое решение относится к системам управления доступом пользователей в Интернет и защита пользователей от нежелательного контента в Интернете.

УРОВЕНЬ ТЕХНИКИ

В данный момент, существует множество способов предотвратить посещение детьми нежелательных сайтов, довольно большое количество программ, ограничивающих к ним доступ. В зависимости от степени владения родителями компьютером, можно найти как более сложные и детальные, так и более простые способы оградить ребенка от нежелательного контента и других рисков Сети, не ограничивая при этом его доступ к огромному количеству полезных образовательных и развлекательных ресурсов в Интернете.

Из патентной заявки US 20150180746 A1 (МПК H04L 12/26, опубл. 25.06.2015) известна система мониторинга и контроля доступа за мобильными устройствами, которая основана том, что на устройство ребенка устанавливается программный модуль (агент), осуществляющий перехват различных данных (данные контактов по телефонной связи, SMS, данные об использовании приложений и веб-сайтов, геолокацию устройства и др., а также по голосовых данных и изображений с видеокамеры). Эти данные передаются с устройства на специализированный сервер, где производится анализ этих данных и принимается решение о доступе ребенка к данному виду информации или о ее блокировке в соответствии с настройками, сделанными родителями. Решения, принятые на сервере, передаются на устройство ребенка.

Недостатком описанной выше системы является то, что требуется обязательная установка на устройство ребенка агента, выполняющего сбор информации. Этот агент должен обеспечить корректное встраивание во все контролируемые приложения. Это позволяет использовать данную систему только для ограниченного числа сценариев, например, для контроля за использованием устройства у ребенка, которому родитель может принудительно установить подобное ПО на телефон, планшет или компьютер.

В качестве решений на текущем уровне техники, локально решающих вопросы ограничения доступа пользователей к определенным веб-сайтам с помощью родительского контроля можно рассматривать такие патентные документы и публикации, например, раскрытые в следующих документах: US 20130040629 A1, US 20090213001 A1, RU 2446460 C1, RU 2336561 C2.

Таким образом, известные из уровня техники решения, предназначенные для ограничения доступа пользователей к определенным веб-сайтам с помощью родительского контроля, имеют ограниченную функциональность, поскольку в них не предусмотрена возможность определения тематики ресурсов непосредственно по тексту веб-страницы, которую запросил пользователь, с учетом лингвистических особенностей языка, на котором написана данная веб-страница, а также имеют недостаточную надежность защиты пользователей от нежелательного контента из сети.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Технической проблемой, на решение которой направлено заявленное техническое решение, является создание способа использования системы определения тематики документов для целей информационной безопасности, который с помощью системы автоматического определения тематик документов, размещенных на веб-страницах, запрошенных пользователем в сети, позволяет в режиме реального времени осуществлять автоматизированную обработку и анализ содержания запрашиваемых пользователем веб-страниц, чтобы защитить от просмотра потенциально вредных и

нежелательных материалов, а именно блокировать неподходящие и опасные веб-сайты.

Технический результат заключается в надежной защите пользователей от нежелательного содержания запрашиваемых веб-страниц, за счет автоматического определения тематик веб-страниц.

5 В предпочтительном варианте реализации заявлен способ использования системы определения тематики документов для целей информационной безопасности, заключающийся в выполнении этапов, на которых:

с помощью системы автоматического определения тематик документов, размещенных на веб-страницах, запрошенных пользователем в сети, в состав которой входят
10 устройства пользователей, внешний сервер и фильтрующий сервер:

формируют запрос на доступ к веб-странице в сети с устройств пользователей;

внешний сервер получает запрос пользователя, формирует ответ и отправляет его обратно к пользователю;

15 фильтрующий сервер перехватывает сформированный ответ для пользователя и с помощью модуля автоматического определения тематик документа осуществляет извлечение текстовой части из полученной веб-страницы, анализирует извлеченный текст запрошенного документа и разбивает текст на уровень предложений и уровень слов;

осуществляет лемматизацию полученных предложений и слов;

20 сравнивает приведенные предложения и слова к словарной форме с базой классифицированных нежелательных слов, словосочетаний и цитат;

на основе полученного сравнения осуществляет автоматическое определение тематик проанализированного документа и формирует список тематик, к которым относится этот документ;

25 полученный набор тематик документа сравнивается с профилем доступа пользователя, в котором зафиксированы критерии, определяющие, может ли пользователь получить доступ к документу, отнесенному к данному набору тематик или нет;

30 по результатам сравнения блокируют или разрешают доступ к документу на веб-странице.

В частном варианте извлеченную текстовую часть разбивают на уровень предложений и уровень слов, используя модули морфологии.

В частном варианте сеть может быть локальной или глобальной.

35 В частном варианте автоматически распознают язык текстовой части полученной веб-страницы.

ОПИСАНИЕ ЧЕРТЕЖЕЙ

Реализация изобретения будет описана в дальнейшем в соответствии с прилагаемыми чертежами, которые представлены для пояснения сути изобретения и никоим образом не ограничивают область изобретения. К заявке прилагаются следующие чертежи:

40 Фиг. 1 иллюстрирует общий вид системы автоматического определения тематик документов;

Фиг. 2 иллюстрирует блок-схему выполнения способа;

Фиг. 3 иллюстрирует блок-схему процедуры автоматического определения тематики документа;

45 Фиг. 4 иллюстрирует пример выполнения вычислительного устройства.

ДЕТАЛЬНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

В приведенном ниже подробном описании реализации изобретения приведены многочисленные детали реализации, призванные обеспечить отчетливое понимание

настоящего изобретения. Однако, квалифицированному в предметной области специалисту, будет очевидно каким образом можно использовать настоящее изобретение, как с данными деталями реализации, так и без них. В других случаях хорошо известные методы, процедуры и компоненты не были описаны подробно,

чтобы не затруднять излишне понимание особенностей настоящего изобретения. Кроме того, из приведенного изложения будет ясно, что изобретение не ограничивается приведенной реализацией. Многочисленные возможные модификации, изменения, вариации и замены, сохраняющие суть и форму настоящего изобретения, будут очевидными для квалифицированных в предметной области специалистов.

Настоящее изобретение направлено на обеспечение способа использования системы определения тематики документов для целей информационной безопасности.

В заявленном техническом решении автоматическое определение тематик документов, запрошенных веб-страниц, осуществляется с использованием базы классифицированных образцов терминов, словосочетаний и цитат, учитывающей лингвистические особенности текста на естественном языке. Предложенная процедура позволяет принимать решение о доступе непосредственно по тексту документа, в том числе без привлечения информации о параметрах, не связанных с содержимым документа. Это позволяет принимать правильное решение о предоставлении доступа даже в случае перемещения документа, изменения содержания документа, в том числе в случаях динамического содержания или высокой степени персонализации документа.

Как представлено на Фиг. 1, система автоматического определения тематик документов (100) включает устройства пользователей (110), фильтрующий сервер (120) и внешний сервер (130).

В качестве устройств пользователей (110) может использоваться смартфон, планшет, персональный компьютер, ноутбук, игровая приставка, телевизор с функцией Smart-TV и т.п.

Пользовательское взаимодействие с внешними серверами в публичных компьютерных сетях осуществляется с помощью программных средств доступа к веб-сервисам - браузеров или приложений, взаимодействующих с удаленными веб-серверами по стандартным протоколам взаимодействия с веб-серверами. Для нормальной работы приложений не требуется дополнительная модификация программных средств доступа к веб-сервисам или установка дополнительных программных агентов на устройстве пользователя. Взаимодействие с этими приложениями осуществляется с помощью стандартных средств ввода/вывода (средства В/В), обеспечивающих распознавание команд с последующим их преобразованием в машиночитаемую форму и вывод информации пользователю.

К таким средствам В/В можно отнести, не ограничиваясь, клавиатуру, сенсорный дисплей, экран, монитор, динамики, проектор, пульт Д/У, манипулятор, мышь, джойстик, трекбол, тачпад, световое перо, стилус и т.п.

Предлагаемый способ использования системы определения тематики документов для целей информационной безопасности предполагает выполнение следующей последовательности операций (см. фиг. 2).

На первом этапе пользователь на своем устройстве формирует запрос на доступ к веб-странице в сети (202).

Внешний сервер с помощью программного обеспечения обрабатывает данный запрос, формирует ответ, содержащий запрошенную пользователем веб-страницу, и передает его обратно на пользовательское устройство (203). При этом внешний сервер может быть размещен, как в локальной сети, так и глобальной сети Интернет.

Ответ, сформированный внешним сервером, перехватывает фильтрующий сервер (204). На следующем этапе, фильтрующий сервер с помощью модуля автоматического определения тематик документа осуществляет извлечение текстовой части из полученной веб-страницы, анализирует извлеченный текст запрошенного документа и разбивает
 5 текст на уровень предложений и уровень слов (205). Осуществляет лемматизацию полученных предложений и слов (206). Сравнивает (207) приведенные предложения и слова к словарной форме с базой классификации (306, см. фиг. 3), в которой хранятся нежелательные слова, словосочетания и цитаты.

На основе полученного сравнения фильтрующий сервер с помощью модуля
 10 автоматического определения тематик документа осуществляет автоматическое определение тематик проанализированного документа и формирует список тематик, к которым относится этот документ (208).

На следующем этапе, фильтрующий сервер с помощью блока проверки тематик по профилю доступа пользователя выполняет сравнение списка тематик (209),
 15 распознанных модулем автоматического определения тематики, и сравнивает их с правилами доступа для данного пользователя, определенными в базе профиля доступа. По результатам сравнения формируется решение о том, можно ли пользователю разрешить доступ к запрошенному документу (210) или доступ к нему должен быть заблокирован (211).

Если критерии, определенные в профиле доступа, позволяют пользователю
 20 просмотреть документы распознанных тематик, формируется решение о предоставлении пользователю доступ к документу. В случае если распознанный набор тематик не соответствует критериям профиля доступа, формируется решение о блокировке доступа. В этом случае доступ к документу блокируется.

Блок проверки тематик может функционировать совместно с системой идентификации
 25 и/или авторизации пользователей (внешней или внутренней), что позволяет в частных реализациях формировать различные правила фильтрации для разных групп пользователей. Система авторизации может поддерживать различные виды идентификации пользователей. В частных реализациях могут использоваться авторизации
 30 пользователя по логину и паролю, авторизация по аппаратному ключу или смарт-карте, идентификация пользователей по их адресу, точке подключения, геолокации и т.д.

В ряде случаев (если это определено в профиле доступа) фильтрующий сервер может
 35 выполнить частичную блокировку материалов документа. При обнаружении в документе блоков текста, не соответствующих правилам доступа, данные блоки текста могут быть удалены из документа или заменены на последовательность знаков, замещающих собой нежелательную часть текста. При этом остальные части документа, не содержащие недопустимых материалов, и их форматирования сохраняются.

В зависимости от тематики запрошенного документа ответ для пользователя может
 40 содержать исходный документ (если документ соответствует профилю доступа пользователя), его модифицированную копию (если профилю не соответствуют только части запрошенного документа) или сообщение о запрещении доступа к запрошенному документу (в случае, если тематика документа не соответствует профилю доступа пользователя).

Структура и логика работы автоматического определения тематики страницы
 45 подробнее рассматривается на Фиг. 3. Фильтрующий сервер на вход (301) блока автоматического определения тематики страницы направляет перехваченный ответ для пользователя, где с помощью модуля выделения текста извлекают и выделяют текстовую часть документа из полученной веб-страницы (302). Данный модуль

анализирует полученное в электронном виде представление документа и определяет его формат: например, документ в формате HyperText Markup Language (HTML), в формате офисного документа или структурированный ответ веб-сервера в формате JavaScript Object Notation (JSON) и т.п. В зависимости от распознанного типа документа

5 вызывается подходящий Модуль разбора соответствующего формата.

Результатом работы является текстовое представление документа с сохранением естественного разделения текста на структурные единицы (разделы, абзацы и т.п.), а также информация о выделении фрагментов текста (жирный шрифт, курсив, цветное выделение и т.п.). Дальнейшая работа производится с этим текстовым представлением.

10 Следующим этапом обработки является разделение текста на отдельные слова и предложения (303). При выделении слов используются Модули морфологии, позволяющие учесть лингвистические особенности словообразования в тексте, а также определить параметры словоформы (склонение, падеж, число), в которой в тексте употребляется данное слово. Параллельно производится синтаксический разбор текста

15 и в тексте выделяются отдельные предложения. Результатом данной стадии является последовательность отдельных слов, при которых сохранена информация о положении их в тексте (координатная информация).

Далее производится лемматизация текста (304), при которой текст преобразуется в последовательность числовых идентификаторов, каждый из которых однозначно

20 идентифицирует нормальную форму каждого слова в тексте. Для этой операции также используются Модули морфологии. При каждом идентификаторе также сохраняется информация о положении слова в тексте и предложении, информация о форме, в которой использовано данное слово и информация о выделении слова. Данная информация используется для учета связей внутри предложения, используемой при расчете степени

25 соответствия тематики анализируемому тексту.

Полученный лемматизированный текст используется для определения тематики документа.

Для определения тематики в цифровом образе текста производится поиск слов, словосочетаний и цитат (305), включенных в базу классификации (306). Термины базы

30 классификации предварительно также лемматизируются.

Для классификации текста производится поиск всех терминов базы классификации в обрабатываемом тексте. При этом производится поиск не только точных вхождений, но также вхождений отдельных слов терминов в анализируемый текст. В результате поиска формируется список слов для каждого из терминов, содержащихся в базе

35 классификации, с данными о положении этих слов в тексте.

После этого производится оценка вероятности соответствия текста тематикам (307), представленным в базе классификации (306). При оценке вероятности учитывается наличие в анализируемом тексте точного или частичного появления текста термина из базы классификации, взаимное расположение найденных слов термина в предложении

40 и тексте в целом, присутствие слов термина в одном предложении, частота встречаемости терминов, форма и согласование слов в найденных терминах, наличие выделения в тексте слов, входящих в термин, наличие в тексте других терминов из данной тематики или других близких тематик. Результатом данной процедуры является список тематик с оценками вероятности соответствия анализируемого текста заданной тематике.

45 Из списка тематик отбирается набор тематик, которые наиболее близко соответствуют данному тексту. Из них формируется финальный список тематик, к которым относится данный текст (308). Данный список тематик используется в качестве результата автоматического определения тематики документа.

Набор тематик, распознаваемых системой, и правила классификации обрабатываемых текстов определяется набором слов, словосочетаний и цитат, включенных в базу классификации (306). Это позволяет настраивать систему классификации на требуемый набор тематик. Ниже приведен пример тематик, содержание которых является
 5 нежелательными для просмотра несовершеннолетними и доступ этих пользователей к документам таких тематик должен быть заблокирован:

- Порнография - материалы порнографического и эротического характера.

Распространение таких материалов среди детей запрещено законодательно;

- Контент для взрослых - к данной категории прежде всего относятся материалы,
 10 эксплуатирующие интерес к сексуальной жизни, но при этом тематически, как правило, не принадлежащие к порнографии или эротике;

- Знакомства - знакомства в Интернете, в том числе знакомства для взрослых, интим-знакомства, услуги по организации знакомств;

- Нецензурная лексика - нецензурная и грубая лексика, в том числе шутки, анекдоты,
 15 в том числе художественные произведения, содержащие нецензурную лексику;

- Наркотики - информация о наркотиках, в том числе рецепты приготовления и употребления, о неправильном использовании лекарственных препаратов, новости о преступлениях, связанных с наркотиками, а также материалы о вреде наркотиков, о борьбе с наркоманией и т.п.;

- Алкоголь - пропаганда алкоголя, рецепты и способы приготовления и употребления
 20 алкогольных напитков, новости о происшествиях с участием лиц, находящихся в состоянии алкогольного опьянения;

- Пропаганда курения - пропаганда курения, рецепты приготовления и способы употребления табачных изделий, материалы о борьбе с курением;

- Азартные игры - азартные игры и лотереи, игры на реальные или виртуальные
 25 деньги, информация и рекомендации по размещению ставок, участию в лотереях, азартных играх, виртуальные казино и рискованные предприятия, спортивные пари и тотализаторы;

- Самоубийства - Информация о самоубийствах, включая новости, пропаганду,
 30 способы, истории и т.п., которая потенциально может побудить детей к совершению действий, представляющих угрозу их жизни;

- Насилие, жестокость - информация о насильственных действиях против человека или животных, насильственная преступность, сексуальное насилие, жестокие компьютерные игры, новости о насилии, а также материалы, способные вызвать страх,
 35 ужас или панику;

- Экстремизм, национализм - категория, прежде всего, предназначена для блокирования экстремистских материалов, распространение которых на территории Российской Федерации запрещено;

- Взрывчатые вещества, оружие - материалы с информацией об изготовлении
 40 взрывчатых веществ и оружия; и в первую очередь - о способах их изготовления в домашних условиях;

- Секты/эзотерика - религиозные секты, сайты сект и информация о сектах, в том числе псевдохристианские, псевдовосточные, неоязыческие секты, психокульти и др. В Методических рекомендациях Министерства образования информация такого рода
 45 названа не совместимой с задачами образования;

- Учебные работы - готовые домашние задания, шпаргалки, рефераты, курсовые и другие виды учебных работ. Категория блокирует материалы, которые дают возможность учащимся списать или купить готовые учебные работы вместо того,

чтобы выполнить их самостоятельно.

На Фиг. 4 представлен пример исполнения вычислительного устройства (400) для реализации функционала по управлению доступом пользователей в Интернет и защита пользователей от нежелательного контента в Интернете.

5 В общем случае устройство (400) содержит такие компоненты, как: один или более процессоров (401), по меньшей мере один блок оперативной памяти (402), средство хранения данных (403), интерфейсы ввода/вывода (404), средство В/В (405), средство сетевого взаимодействия (406), универсальную шину (410).

10 Процессор (401) устройства выполняет основные вычислительные операции, необходимые для функционирования системы (100) или функционала одного или более ее компонентов. Процессор (401) исполняет необходимые машиночитаемые команды, содержащиеся в оперативной памяти (402).

Память (402), как правило, выполнена в виде ОЗУ и содержит необходимую программную логику, обеспечивающую требуемый функционал.

15 Средство хранения данных (403) может выполняться в виде HDD, SSD дисков, рейд массива, флэш-памяти, оптических накопителей информации (CD, DVD, MD, Blue-Ray дисков) и т.п. Средства (403) позволяют выполнять долгосрочное хранение различного вида информации, например, истории обработки поисковых запросов, идентификаторов пользователей и т.п.

20 Интерфейсы (404) представляют собой стандартные средства для подключения и работы с серверами, например, USB, RS232, RJ45, LPT, COM, HDMI, PS/2, Lightning, Fire Wire и т.п.

Выбор интерфейсов (404) зависит от конкретного исполнения устройства (400), которое может представлять собой персональный компьютер, мейнфрейм, серверный 25 кластер, тонкий клиент, смартфон, ноутбук и т.п.

В качестве средств В/В данных (405) может использоваться: клавиатура, джойстик, дисплей (сенсорный дисплей), проектор, тачпад, манипулятор мышь, трекбол, световое перо, динамики, микрофон и т.п.

30 Компоненты системы соединяются посредством средств сетевого взаимодействия (406), которые выбираются из средств, предназначенных для реализации процесса коммуникации между различными блоками посредством проводной и/или беспроводной связи, в частности, такими типами связи могут являться: GSM, GPRS, LTE, 5G, WLAN, WMAN, Wi-Fi, Bluetooth, BLE, спутниковая связь, NFC, Ethernet, USB, IrDa, Lightning, UART, CAN и т.п.

35 Взаимодействие всех компонентов устройства производится путем обмена данными через универсальную шину (410).

40 Система управления доступом сохраняются на Средстве хранения данных (403) в виде последовательности команд и загружаются в Память (402) при запуске устройства. Процессор (401) выполняет последовательность команд системы управления доступом и осуществляет взаимодействие с пользователями системы и внешними серверами посредством интерфейсов (406). Взаимодействие с администратором системы, выполняющим настройку и управление работой системы управления доступом, осуществляется через средства В/В данных (404 и 405).

45 Данные, необходимые для работы системы управления доступом, могут быть сохранены на Средстве хранения данных (403) или к ним может быть предоставлен удаленный доступ через набор интерфейсов (406).

В настоящих материалах заявки было представлено предпочтительное раскрытие осуществление заявленного технического решения, которое не должно использоваться

как ограничивающее иные, частные воплощения его реализации, которые не выходят за рамки испрашиваемого объема правовой охраны и являются очевидными для специалистов в соответствующей области техники.

(57) Формула изобретения

1. Способ автоматического управления доступом к веб-странице, заключающийся в выполнении этапов, на которых:

с помощью системы автоматического определения тематик документов, размещенных на веб-страницах, запрошенных пользователем в сети, в состав которой входят

устройства пользователей, внешний сервер и фильтрующий сервер:

формируют запрос на доступ к веб-странице в сети с устройств пользователей;

внешний сервер получает запрос пользователя, формирует ответ и отправляет его обратно к пользователю;

фильтрующий сервер перехватывает сформированный ответ для пользователя и с помощью модуля автоматического определения тематик документа осуществляет извлечение текстовой части из полученной веб-страницы, анализирует извлеченный текст запрошенного документа и разбивает текст на уровень предложений и уровень слов;

осуществляет лемматизацию полученных предложений и слов;

сравнивает приведенные предложения и слова к словарной форме с базой классифицированных нежелательных слов, словосочетаний и цитат;

на основе полученного сравнения осуществляет автоматическое определение тематик проанализированного документа и формирует список тематик, к которым относится этот документ;

полученный набор тематик документа сравнивается с профилем доступа пользователя, в котором зафиксированы критерии, определяющие, может ли пользователь получить доступ к документу, отнесенному к данному набору тематик, или нет;

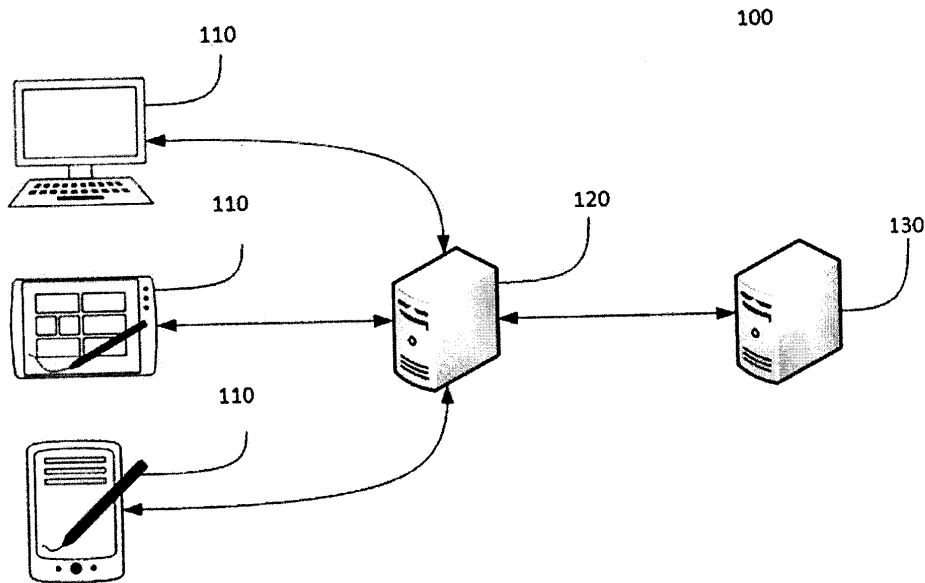
по результатам сравнения блокируют или разрешают доступ к документу на веб-странице.

2. Способ по п. 1, характеризующийся тем, что извлеченную текстовую часть разбивают на уровень предложений и уровень слов, используя модули морфологии.

3. Способ по п. 1, характеризующийся тем, что сеть может быть локальной или глобальной.

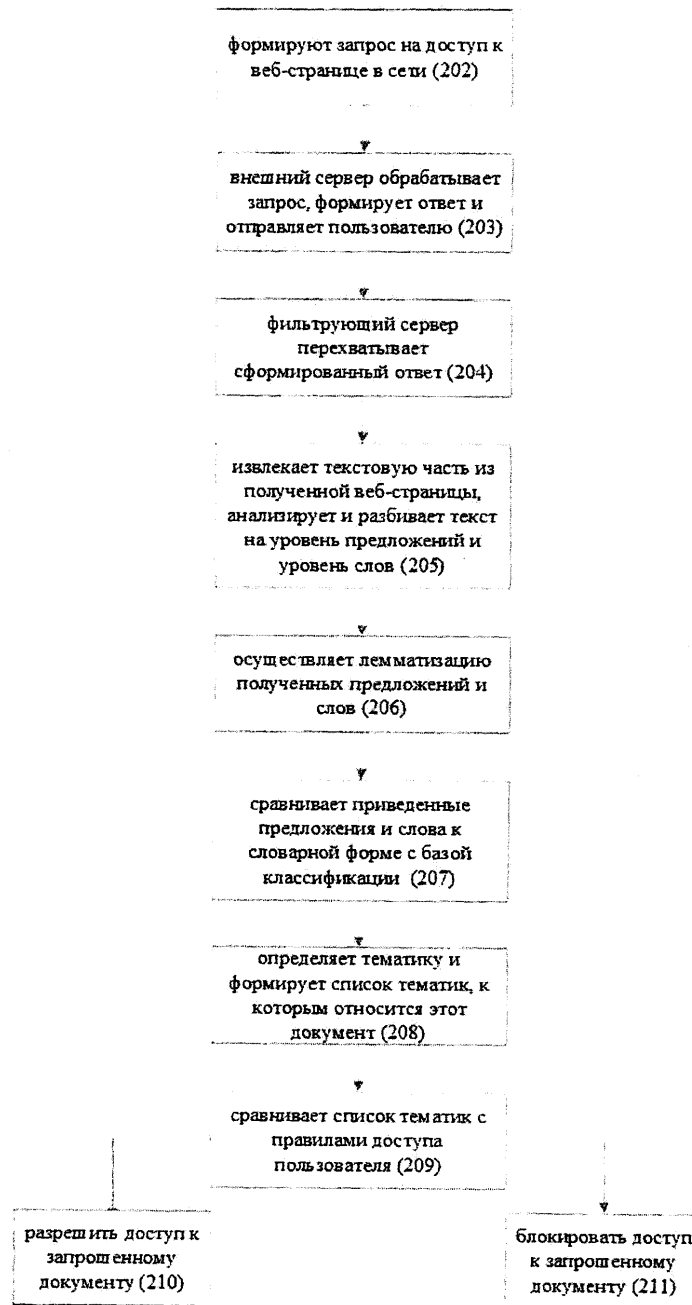
4. Способ по п. 1, характеризующийся тем, что автоматически распознают язык текстовой части полученной веб-страницы.

1

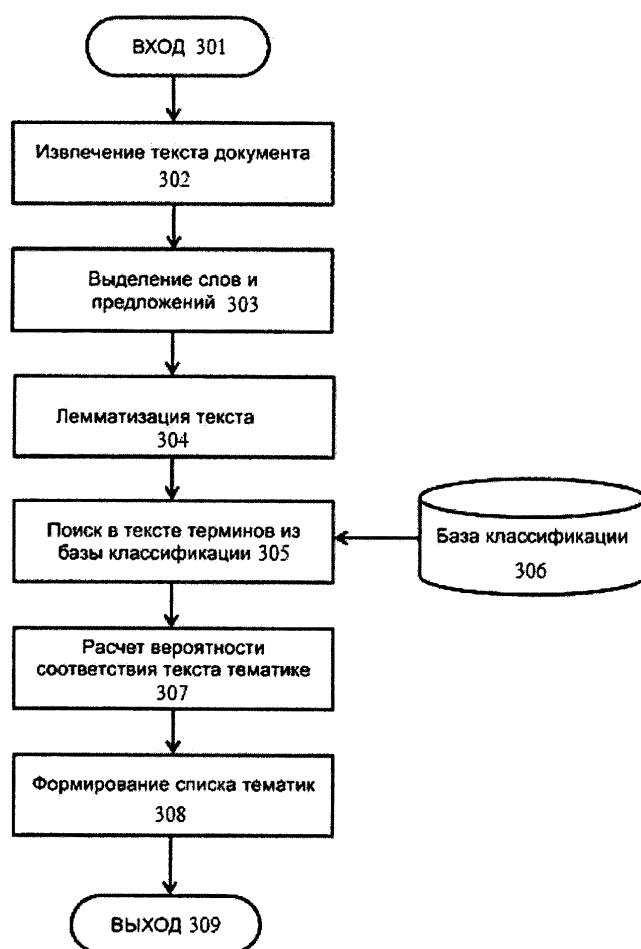


Фиг. 1

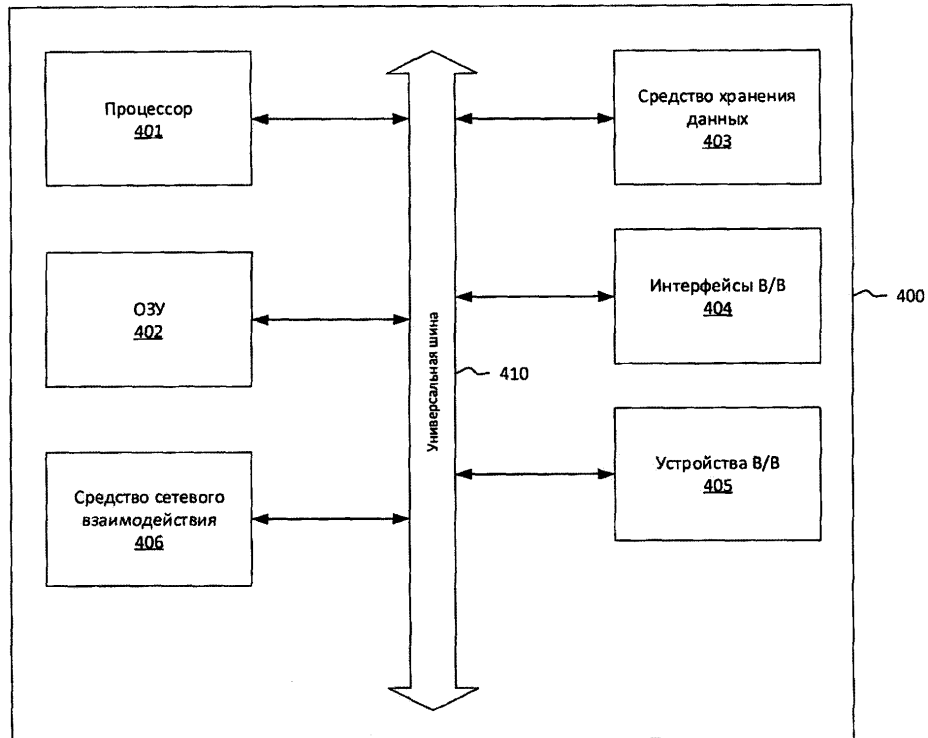
2



Фиг. 2



Фиг. 3



Фиг. 4