

(19)



**Евразийское
патентное
ведомство**

(21) **202200059** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки
2023.03.31

(51) Int. Cl. **G06F 17/00** (2019.01)
G06F 16/21 (2019.01)
G06F 16/9538 (2019.01)

(22) Дата подачи заявки
2021.09.21

(54) **ДОМЕННО-ОРИЕНТИРОВАННАЯ ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА
(ДИПС) (СПОСОБ ЕЕ СОЗДАНИЯ)**

(96) **2021/025 (AZ) 2021.09.21**

(72) Изобретатель:

(71) Заявитель:
**ИНСТИТУТ СИСТЕМ
УПРАВЛЕНИЯ НАНА (AZ)**

**Аббасов Али Мамед оглы, Касумов
Вагиф Алиджавад оглы (AZ)**

(57) Изобретение относится к информационно-поисковым системам. Сущность изобретения состоит в создании системы для оптимизации информационного поиска в интернете, позволяющей относительно быстро обнаружить наиболее релевантные результаты. Технический эффект заявляемого изобретения позволяет повысить уровень эффективности информационного поиска путем систематизации Web-пространства, сужения области поиска и организации тематически направленного поиска, оставляя при этом без изменения программно-технические параметры поисковых систем.

A1

202200059

202200059

A1

**ДОМЕННО-ОРИЕНТИРОВАННАЯ ИНФОРМАЦИОННО-
ПОИСКОВАЯ СИСТЕМА (ДИПС)
(СПОСОБ ЕЕ СОЗДАНИЯ)**

Изобретение относится к информационно-поисковым системам.

Известно, что пользователь, выполняя веб-поиск с использованием поисковой системы, преследует две основных цели. Ему нужно, чтобы поисковая система выдавала наиболее релевантные результаты, и чтобы это было выполнено относительно быстро. Однако анализ проблем информационного пространства Интернет, а также исследований в области поиска информации показывает, что в современных условиях астрономического роста объема информации в среде Интернет эффективность средств поиска информации явно отстает от желаемого уровня. Главной причиной данной проблемы, в основном, является неорганизованность информационных ресурсов, в том числе Web-документов в среде Интернет. Данное обстоятельство связано следующими факторами:

- информационные ресурсы, в том числе Web-документы Интернет создаются независимо друг от друга в произвольной форме, по желанию авторов или владельцев;
- каждый информационный ресурс характеризуется только определенным набором признаков, в лучшем случае, и их частотными характеристиками;
- ссылки между информационными ресурсами ставятся без учета их содержания, в целом, в результате которого логическая связь между информационными ресурсами существует только между соседними, в лучшем случае через две-три ссылки;
- существующие поисковые системы, в основном, дают возможность квазислучайного (малонаправленного) поиска по паутине Web,

эффективность которого зависит от библиотечной подготовленности пользователей.

В настоящее время для решения указанной проблемы предлагаются различные способы и системы отбора информации в сети Интернет.

Известна система и способ (1) для получения информации от распределенного с использованием поискового агента, работающей в сети, по «всемирной паутине». Система предназначена для обеспечения клиентов сервера быстрыми и точными данными поиска веб-агентом и основана на анализе данных, полученной информации от работающих в сети систем с использованием распределенного поискового веб-агента. Собранный поисковым веб-агентом сервера информацию сравнивают с данными полученной информации, распространенной через страницу результатов поискового механизма, либо посредством клиентских словарей, исходящих от сервера, которые обобщают данные поискового веб-агента. Предполагается что техническим результатом является обеспечение клиентов сервера быстрыми и точными данными поиска веб-агентом.

Известен способ (2) отбора информации в сети интернет и использования этой информации в разделяемом веб-сайте, имеющем некоторое доменное имя, согласно которому с помощью интернет-робота осуществляют поиск информации в сети Интернет о субъектах. Используемый критерий поиска в виде списка ключевых слов, соответствующих доменному имени разделяемого веб-сайта, позволяет отобразить информацию, включающую средства индивидуализации субъектов, содержащие в себе соответствующую указанному доменному имени символьную последовательность. Указанный способ не всегда может обеспечить корректный отбор информации из-за того, что соответствие информации документу оценивается путем интернет-робота, что не исключает ошибок в определении тематики документа.

Задача изобретения состоит в создании поисковой системы для оптимизации информационного поиска в интернете, позволяющей относительно быстро обнаружить наиболее релевантные результаты.

Сущность изобретения состоит в способе создания доменно-ориентированной информационно-поисковой системы (ДИПС). Способ, предложенный в рамках информационной модели доменов (IM_D) включает в себя разделение виртуального информационного пространства на семантически слабосвязанные домены, являющиеся зонами охвата отдельных поисковых систем, создание многоуровневой иерархической структуры, на каждом уровне которой формируются доменные области со своими атрибутами, имеющие предметно-содержательную связь, формирование пользовательского запроса (UQ), организацию поиска информации с использованием модулей «МАРР2» (Mapping 2), «МАРР1» (Mapping 1), оценки релевантности и ранжирование результатов.

Информационная модель доменов (IM_D) представляет собой информационное пространство, организованное как иерархическая структура, на уровнях которой определяются Web-области, содержащие объекты со своими атрибутами и состоящие из доменных областей, включающие в себя множества объектов доменов, в том числе серверов Web-сайтов и других информационных сервисов. Каждый объект описывается множеством атрибутов, дескрипторов, ключевых слов и терминов. Определяются принципы размещения объектов на уровнях, отношения между объектами, между объектами и их атрибутами, а также между атрибутами объектов.

Предложенная для описания ресурсов информационного пространства Интернета информационная модель доменов (IM_D) позволяет также построить модель виртуального Web-окружения (Virtual Web Environment – VWE) и модель пользовательского запроса (User Query - UQ), а также разработать методы отображения одной модели в другую, определения релевантности отображения и ранжирования результатов.

Информационную модель доменов (IM_D) в общем виде можно представить в следующем виде:

$$IM_D \rightarrow \{N, E, H(E), H(E, E^*), A, R(E, A), R(A, A^*), R(E, E^*)\},$$

где

- N – максимальное количество уровней в иерархии доменов Web-пространства;
- $E = \{e_i\}_n$ – множество объектов (сущностей) в домене;
- $A = \{a_j\}_m$ – множество атрибутов, используемых для описания объектов домена;
- $H(E)$ – матрица (или вектор) размещения объектов (категорий) на уровнях иерархии доменов, где $H(e_i) = 1$, если объект e_i размещен на домене, $H(e_i) = 0$, в обратном случае, $i = \overline{1, n}$;
- $H(E, E^*)$ – матрица отношений категоризации (вложенности на subtype и supertype) между объектами, где $H(e_i, e_j^*) = 1$, если объект e_i имеет отношение вложенности на subtype или supertype с объектом e_j^* , $H(e_i, e_j^*) = 0$, в обратном случае, $i, j = \overline{1, n}$;
- $R(E, A)$ – матрица отношений между объектами и атрибутами доменов, где $R(e_i, a_j) = 1$, если объект e_i имеет отношение релевантности с атрибутом a_j , $R(e_i, a_j) = 0$, в обратном случае, $i = \overline{1, n}, j = \overline{1, m}$;
- $R(E, E^*)$ – матрица отношений между объектами доменов где $R(e_i, e_j^*) = 1$, если объект e_i имеет отношение с объектом e_j^* , $R(e_i, e_j^*) = 0$, в обратном случае, $i, j = \overline{1, n}$.

Исходя из вышеприведенной модели IM_D , информационную модель виртуального Web-окружения (VWE) можно представить как:

$$VWE \rightarrow \{URL, K, R(URL, K), R(K, K), R(URL, URL)\},$$

где

- URL – множество связанных (релевантных) Web документов Интернета;
- K – множество дескрипторов (ключевых слов), используемые для описания Web документов;

- $R(URL, K)$ – матрица отношений между Web документами и ключевыми словами;
- $R(K, K)$ – матрица взаимосвязей между ключевыми словами внутри Web документов;
- $R(URL, URL)$ – матрица связей между Web документами.

VWE является частью Web окружения Интернета и относится (релевантна) рассматриваемому конкретному домену. VWE содержит информацию о доменах Интернета форме, подобной IM и использует модель Web-документа.

Множество Web документов URL также является объектами домена, поэтому можно писать $URL \subset E$. Тогда, множество дескрипторов (ключевых слов) K также является подмножеством множества атрибутов, т.е. $K \subset A$. Учитывая это, матриц отношений $R(URL, K)$, $R(K, K)$ и $R(URL, URL)$ можно рассматривать как $R(K^E, K^A)$, $R(K^A, K^A)$ и $R(K^E, K^{E*})$. Здесь K^E – подмножество дескрипторов (ключевых слов), описывающих объект E , K^A – множество дескрипторов (ключевых слов), описывающих Web-документы, т.е. $K^E \in E$ и $K^A \in A$.

Матрицы взаимосвязей модели VWE представляется следующим образом:

- $R(K^E, K^A)$ – определяет Web-документы, аналогично отношениям между объектами и их атрибутами, в виде

$$V_1 = \|v_{ij}^1\|_{n \times m},$$

где

$$v_{ij}^1 = \begin{cases} URL, & \text{если } R(K^E, K^A) = 1 \\ 0, & \text{в обратном случае} \end{cases}$$

- $R(K^E, K^{E*})$ – определяет Web-документы, согласно отношениям между объектами в виде

$$V_2 = \|v_{ij}^2\|_{n \times n},$$

где

$$v_{ij}^2 = \begin{cases} URL, & \text{если } R(K^E, K^E) = 1 \\ 0, & \text{в обратном случае} \end{cases}$$

- $R(K^A, K^A)$ – определяет взаимосвязь между дескрипторами (ключевыми словами) Web-документов в виде

$$V_3 = \|v_{ij}^3\|_{m \times m},$$

где

$$v_{ij}^3 = \begin{cases} 1, & \text{если ключевые слова } i \text{ и } j \text{ имеют взаимосвязь} \\ 0, & \text{в обратном случае} \end{cases}$$

Последняя матрица предоставляет возможность использовать взаимосвязи между ключевыми словами для повышения эффективности результата поиска, путем учета семантической связанности ключевых слов, их синонимов и ассоциативных слов.

Формирование пользовательского запроса UQ дает возможность пользователю выражать свои требования на иерархии информационной модели и используя существующий браузер, отправлять UQ через модуль отображения «МАРР2» (Mapping 2) в модель VWE. Формируя свой запрос в нижеследующем виде пользователь ищет Web-документы из информационного пространства Интернета, которые, как обычно, содержат ключевые слова, связи из информационной модели доменов (IM_D) и виртуального Web-окружения (VWE):

$$UQ \rightarrow \{E, A, R(K^E, K^A), R(K^E, K^{E*})\},$$

где

$$R(K^E, K^A), R(K^E, K^{E*}) \in \{0,1\}.$$

В этом случае веса ключевых слов и связи из других Web-документов могут быть использованы для фильтрации результатов.

Модуль отображения «МАРР1» (Mapping 1) позволяет создать виртуальное Web-окружение (VWE) с помощью существующих Web-браузеров, представляющие возможность получения информации о доменах. Для создания виртуального Web-окружения (VWE) необходим переход от информационной модели виртуального Web-окружения к информационной

модели заданного домена системы.

Учитывая, что множества ключевых слов взяты из множества дескрипторов $K \in E \cup A$, то преобразование модели VWE в модель IM_D мы можно рассматривать как

$$K \in K^E \cup K^A,$$

где $K^E \in K$ и $K^A \in A$. Для заданных объектов не трудно выделять (классифицировать) K^E из E и K^A из A .

Таким образом, отношения между объектами и атрибутами $R(K^E, K^A)$ в множестве ключевых слов для VWE будет

$$R(K^E, K^A) = \begin{cases} 1, \text{ если } R(K^E, K^A) = 1 \\ 1, \text{ если } R(URL(K^E), URL(K^A)) = 1 \\ 0, \text{ иначе} \end{cases}$$

Выражение $R(K^E, K^A) = 1$ означает, что внутри Web-документов существуют связи между K^E и K^A , а выражение $R(URL(K^E), URL(K^A)) = 1$ означает, что существуют связи между Web-документами, содержащие K^E и K^A . Аналогично выражению $R(K^E, K^A)$, определяются отношения между объектами $R(K^E, K^E)$ с учетом их ключевых слов для VWE:

$$R(K^E, K^A) = \begin{cases} 1, \text{ если } R(K^E, K^E) = 1 \\ 1, \text{ если } R(URL(K^E), URL(K^E)) = 1 \\ 0, \text{ иначе} \end{cases}$$

Модуль MAP2 осуществляет процесс поиска информации в VWE на основе UQ, определяя соответствие между их элементами, выполняет механизм отображения между UQ и VWE, реализует оптимизацию просмотра (поиска) результата, определяет какое направление и какой маршрут просмотра будет наилучшим для планирования UQ. Элементами пользовательского запроса и виртуального Web-окружения являются:

- множество объектов пользовательского запроса E отображается на множество ключевых слов (дескрипторов) K^A , описывающих эти объекты VWE, т.е. $E \rightarrow K^E$;

- множество атрибутов A объектов домена отображается на множество ключевых слов (дескрипторов), описывающих эти атрибуты, т.е. $A \rightarrow K^A$;
- так как множества E и A отображаются, соответственно, на множества K^E и K^A , то отношения между объектами и атрибутами, также между самими объектами будут отображаться на следующие отношения:

$$R(E, A) \rightarrow R(K^E, K^A) \quad \text{и} \quad R(E, E^*) \rightarrow R(K^E, K^{E^*}).$$

Согласно этому отображению, выбор (поиск) Web-документов из VWE по пользовательскому запросу реализуется по одной из следующих строк:

№	$E \rightarrow K$	$A \rightarrow K^A$	$R(E, A) \rightarrow R(K^E, K^A)$	$R(E, E^*) \rightarrow R(K^E, K^{E^*})$	Результаты (Web-документы)
1.	T	T	T	T	$v_1 \cap v_2$
2.	T	T	T	F	v_1
3.	T	T	F	T	v_2
4.	T	T	F	F	$v_1 \cup v_2$
5.	T	F	F	T	v_2
6.	F	T	F	F	v_1
7.	F	F	F	F	-

Принципиальная схема ДИПС проиллюстрирована на рисунке 1.

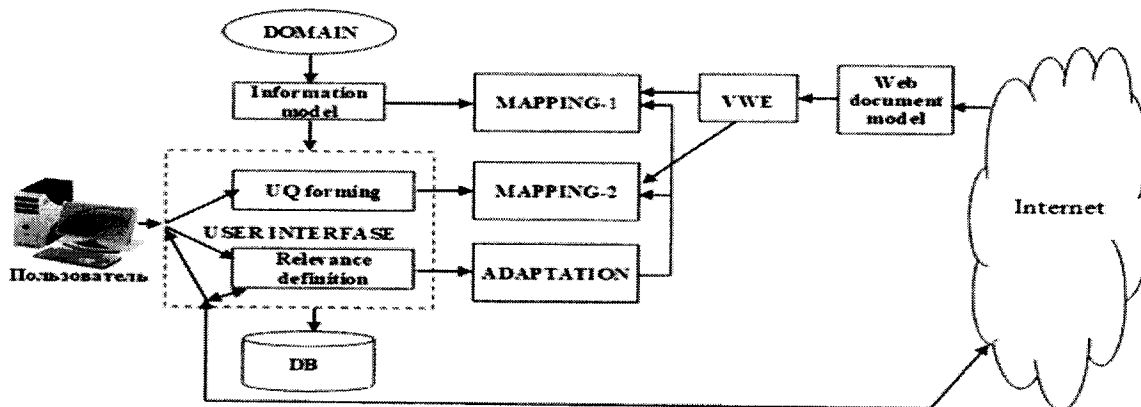


Рисунок 1. Общая структура ДИПС

Структура ДИПС включает в себя модуль пользователя – Web браузер пользователя; пользовательский интерфейс (USER INTERFACE), состоящего из модуля формирования запроса (UQ forming) и модуля определения релевантности (Relevance definition); информационную модель доменов (domain information model) IM_D для описания информационных ресурсов интернета (Internet); модули отображения Mapping-1 и Mapping-2; модель виртуального Web-окружения (Virtual Web Environment – VWE); модель Web-

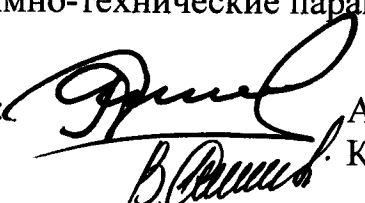
документа (Web document model); модуль адаптации (ADAPTATION); база знаний (DB).

Работа системы осуществляется следующим образом.

Модуль интерфейса пользователя (UI) ДИПС выполняет функции формирования поискового запроса пользователя UQ и определения релевантности документов к пользовательским запросам. Пользователь, используя существующий Web-браузер выражает свои информационные потребности и формирует необходимый поисковый запрос UQ, далее отправляет его через блок отображения «MAPP2» в модуль виртуального Web-окружения (VWE). Блок MAPP2, имеющий знание о UQ и VWE выполняет механизм отображения между UQ и VWE. MAPP2 также реализует оптимизацию просмотра (поиска), используя базу данных и определяя наилучшее направление и маршрут процесса поиска для сформированного UQ. Когда пользователь получает список Web-документов из VWE, модуль оценки релевантности результатов поиска осуществляет ранжирование найденных Web-документов на основе критерий релевантности.

Технический эффект заявляемого изобретения позволяет повысить уровень эффективности информационного поиска в Интернете путем систематизации Web-пространства, сужения области поиска и организации тематически направленного поиска, оставляя при этом без изменения программно-технические параметры поисковых систем.

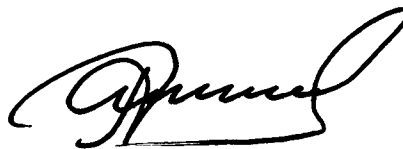
Авторы:



Аббасов Али Мамед оглы

Касумов Вагиф Алиджавад оглы

Генеральный директор Института
Систем Управления НАНА, академик



АББАСОВ А.М.

ЛИТЕРАТУРА

1. Патент РФ № 2383920 «Система и способ для клиент-обоснованного поиска веб-агентом»
2. Патент РФ № 2413278 «Способ отбора информации в сети интернет и использования этой информации в разделяемом веб-сайте и компьютерный сервер для реализации этого способа»

ФОРМУЛА ИЗОБРЕТЕНИЯ

Способ создания доменно-ориентированной информационно-поисковой системы (ДИПС), в рамках предложенной информационной модели доменов (IM_D) включает в себя разделение виртуального информационного пространства на семантически слабосвязанные домены, являющиеся зонами охвата отдельных поисковых систем, создание многослойной иерархической структуры, на каждом уровне которой формируются доменные области со своими атрибутами, имеющие предметно-содержательную связь, формирование пользовательского запроса (UQ), организацию поиска информации с использованием модулей «МАРР2» (Mapping 2), «МАРР1» (Mapping 1), оценки релевантности и ранжирование результатов.



Информационная модель доменов ($IM_D \rightarrow \{N, E, H(E), H(E, E^*), A, R(E, A), R(A, A^*), R(E, E^*)\}$) представляет собой информационное пространство с иерархической структурой, на уровнях которой определяются Web-области, содержащие объекты со своими атрибутами и состоящие из доменных областей, включающий в себя множества объектов доменов. Каждый объект описывается множеством атрибутов, дескрипторов, ключевых слов и терминов. В рамках данной модели определяются принципы размещения объектов на уровнях, отношения между объектами, между объектами и их атрибутами, а также между атрибутами объектов.

На основе информационной модели доменов (IM_D) предложены информационные модели виртуального Web-окружения ($VWE \rightarrow \{URL, K, R(URL, K), R(K, K), R(URL, URL)\}$) и пользовательского запроса ($UQ \rightarrow \{E, A, R(K^E, K^A), R(K^E, K^{E^*})\}$), а также разработаны методы отображения одной модели в другую, определения релевантности отображения и ранжирования результатов.

Разработана функциональная структура ДИПС. Согласно этой структуре модуль отображения «МАРР1» (Mapping 1) позволяет создать виртуальное

Web-окружение (VWE) с помощью существующих Web-браузеров, представляющие возможность получения информации о доменах. Для создания виртуального Web-окружения (VWE) необходим переход от информационной модели виртуального Web-окружения к информационной модели заданного домена системы. Модуль MAPP2 осуществляет процесс поиска информации в VWE на основе UQ, определяя соответствие между элементами пользовательского запроса и виртуального Web-окружения, реализует оптимизацию просмотра (поиска). После выдачи списка Web-документов, найденных из VWE в результате поиска, модуль оценки релевантности осуществляет ранжирование найденных Web-документов на основе критерий релевантности.

Такой подход предоставляет возможность целенаправленно влиять на процесс информационного поиска. Предложенный способ позволяет повысить уровень эффективности информационного поиска путем систематизации Web-пространства, сужения области поиска и организации тематически направленного поиска, оставляя при этом без изменения программно-технических параметров поисковых систем.

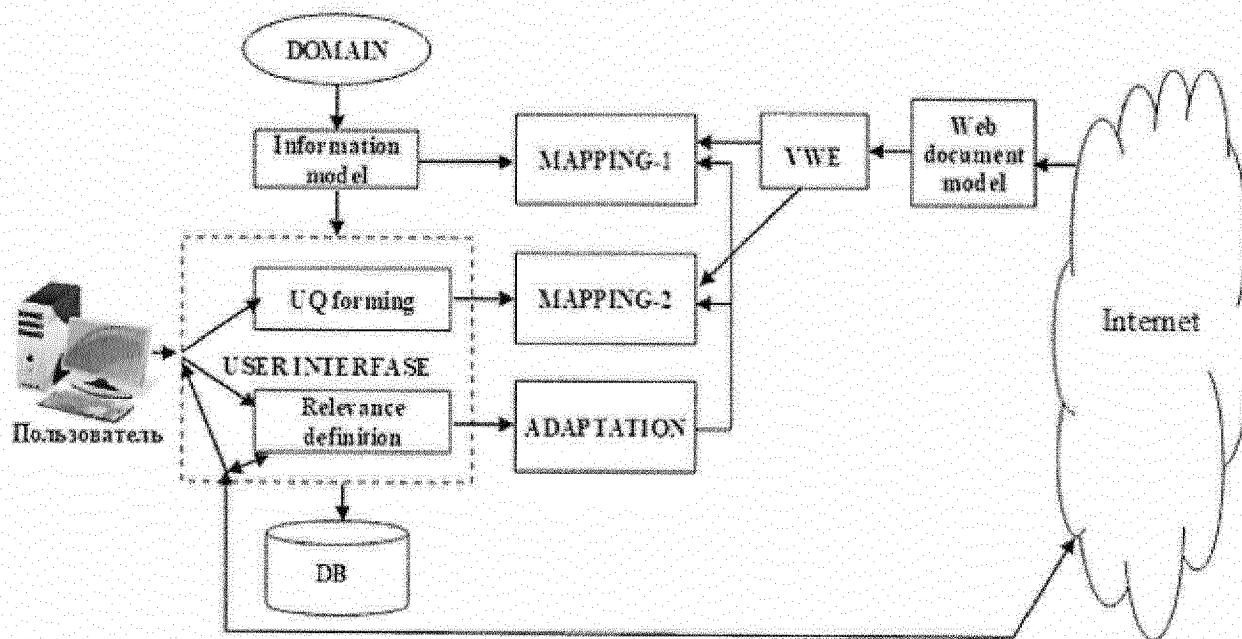
Авторы:  Аббасов Али Мамед оглы
 Касумов Вагиф Алиджавад оглы

Генеральный директор Института
Систем Управления НАНА, академик



АББАСОВ А.М.

ДОМЕННО-ОРИЕНТИРОВАННАЯ ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА (СПОСОБ ЕЕ СОЗДАНИЯ)



Фиг.1

Авторы:

Аббасов Али Мамед оглы

Касумов Вагиф Алиджавад оглы

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ
(статья 15(3) ЕАПК и правило 42 Патентной инструкции к ЕАПК)

Номер евразийской заявки:

202200059

А. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:

G06F 17/00 (2019.01)
G06F 16/21 (2019.01)
G06F 16/9538 (2019.01)

Согласно Международной патентной классификации (МПК)

Б. ОБЛАСТЬ ПОИСКА:

Просмотренная документация (система классификации и индексы МПК)
G06F 7/00, 16/00, 16/90 - 16/9538, 17/00 - 17/27, G06N 5/00 - 5/02, 20/00

Электронная база данных, использовавшаяся при поиске (название базы и, если, возможно, используемые поисковые термины)
ESP@CENET, K-PION, PAJ, USPTO, WIPO, GOOGLE, ИС «ПОИСКОВАЯ ПЛАТФОРМА» (РОСПАТЕНТ)

В. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
X	US10,817,568 B2, (INTERNATIONAL BUSINESS MACHINES CORPORATION), 27.10.2020 реферат, колонка 2, строка 66 – колонка 3, строка 24, п. 1 формулы	1
X	US2008/0091633 A1, (MICROSOFT CORPORATION), 17.04.2008 реферат, абзацы [0007] - [0013], п.п. 1, 16 формулы	1
A	US2002/0129015 A1, (CAUDILL M. et al), 12.09.2002	1
A	US7,299,222 B1, (AOL LLC), 20.11.2007	1

последующие документы указаны в продолжении

* Особые категории ссылочных документов:

«А» - документ, определяющий общий уровень техники

«D» - документ, приведенный в евразийской заявке

«E» - более ранний документ, но опубликованный на дату подачи евразийской заявки или после нее

«O» - документ, относящийся к устному раскрытию, экспонированию и т.д.

«P» - документ, опубликованный до даты подачи евразийской заявки, но после даты испрашиваемого приоритета"

«Т» - более поздний документ, опубликованный после даты приоритета и приведенный для понимания изобретения

«X» - документ, имеющий наиболее близкое отношение к предмету поиска, порочащий новизну или изобретательский уровень, взятый в отдельности

«Y» - документ, имеющий наиболее близкое отношение к предмету поиска, порочащий изобретательский уровень в сочетании с другими документами той же категории

«&» - документ, являющийся патентом-аналогом

«L» - документ, приведенный в других целях

Дата проведения патентного поиска: **09/08/2022**

Уполномоченное лицо:

Заместитель начальника отдела механики,
физики и электротехники



М.Н. Юсупов