



**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ,
ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: **2008110731/08, 20.09.2006**

(24) Дата начала отсчета срока действия патента:
20.09.2006

Приоритет(ы):

(30) Конвенционный приоритет:
21.09.2005 US 11/231,955

(43) Дата публикации заявки: **27.09.2009** Бюл. № 27

(45) Опубликовано: **27.05.2011** Бюл. № 15

(56) Список документов, цитированных в отчете о поиске: **RU 2236699 C1, 20.09.2004. US 2005165753 A1, 28.07.2005. RU 2001128643 A, 20.07.2003. US 6272507 B1, 07.08.2001. US 2002026390 A1, 28.02.2002. US 2003046389 A1, 06.03.2003.**

(85) Дата начала рассмотрения заявки РСТ на национальной фазе: **20.03.2008**

(86) Заявка РСТ:
US 2006/037206 (20.09.2006)

(87) Публикация заявки РСТ:
WO 2007/035919 (29.03.2007)

Адрес для переписки:
**129090, Москва, ул.Б.Спасская, 25, стр.3,
ООО "Юридическая фирма Городисский и
Партнеры", пат.пов. А.В.Мищу, рег.№ 364**

(72) Автор(ы):

**МЕЙЕРЗОН Дмитрий (US),
САРАГОЗА Хьюго (US),
ПЕЛТОНЕН Кайл (US),
ДЕБРЮИН Эндрю (US)**

(73) Патентообладатель(и):

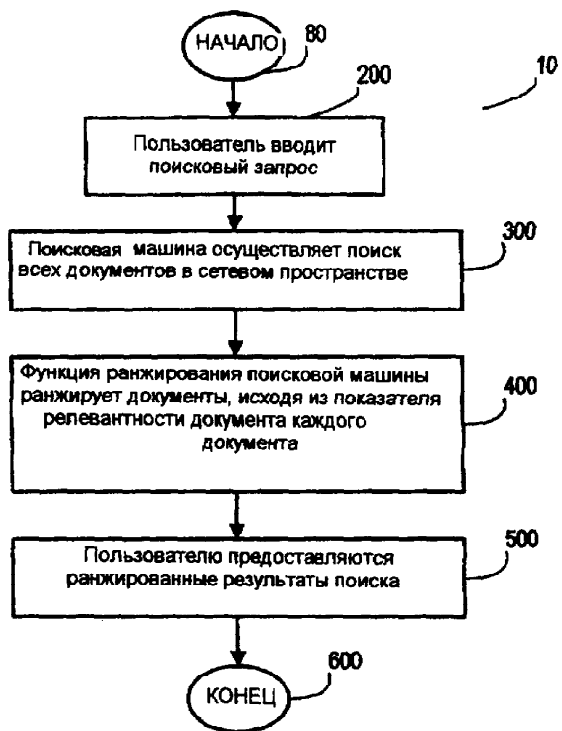
МАЙКРОСОФТ КОРПОРЕЙШН (US)

**(54) ФУНКЦИИ РАНЖИРОВАНИЯ, ИСПОЛЬЗУЮЩИЕ СТАТИСТИЧЕСКИЕ ДАННЫЕ
ИСПОЛЬЗУЕМОСТИ ДОКУМЕНТА**

(57) Реферат:

Изобретение относится к способу ранжирования результатов поиска. Техническим результатом является повышение достоверности результатов поиска. Раскрыты способы предоставления оценки релевантности документа для документа в сети. Также раскрыт машиночитаемый носитель информации, на котором сохранены

исполняемые компьютером инструкции для выполнения способа предоставления оценки релевантности для документа в сети. Помимо этого раскрыты вычислительные способы, содержащие, по меньшей мере, один модуль приложения, который содержит прикладной код для выполнения способов предоставления оценки релевантности для документа в сети. 4 н. и 10 з.п. ф-лы, 4 ил.



ФИГ.1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY,
PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION(21)(22) Application: **2008110731/08, 20.09.2006**(24) Effective date for property rights:
20.09.2006

Priority:

(30) Priority:
21.09.2005 US 11/231,955(43) Application published: **27.09.2009 Bull. 27**(45) Date of publication: **27.05.2011 Bull. 15**(85) Commencement of national phase: **20.03.2008**(86) PCT application:
US 2006/037206 (20.09.2006)(87) PCT publication:
WO 2007/035919 (29.03.2007)

Mail address:

**129090, Moskva, ul.B.Spaskaja, 25, str.3, OOO
"Juridicheskaja firma Gorodisskij i Partnery",
pat.pov. A.V.Mitsu, reg.№ 364**

(72) Inventor(s):

**MEJERZON Dmitrij (US),
SARAGOZA Kh'jugo (US),
PELTONEN Kajl (US),
DEBRJuIN Ehndrju (US)**

(73) Proprietor(s):

MAJKROSOFT KORPOREJShN (US)

(54) RANKING FUNCTIONS USING STATISTICAL DATA OF DOCUMENT USABILITY

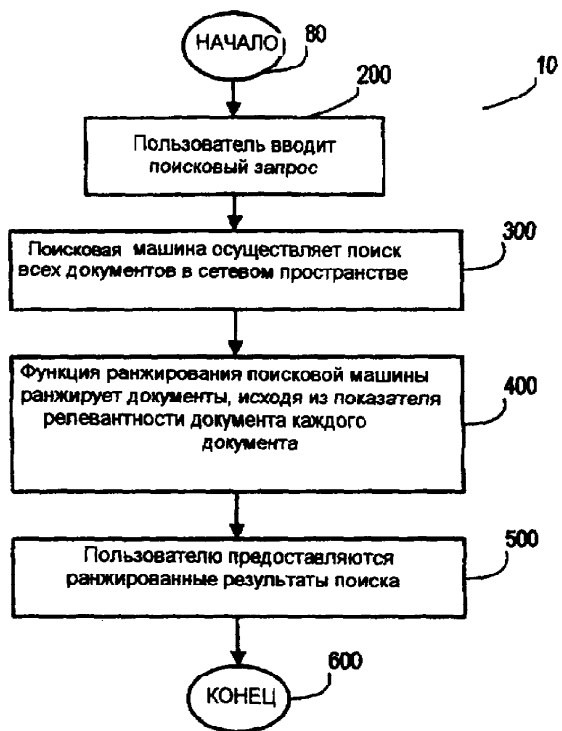
(57) Abstract:

FIELD: information technology.

SUBSTANCE: invention discloses methods of evaluating document relevance in a network. Disclosed also is a machine-readable data medium which stores computer-executable instructions for executing the method of assessing document relevance in a network. The invention also discloses computing methods comprising at least one application module which contains an application code for executing methods of assessing document relevance in a network.

EFFECT: high reliability of search results.

14 cl, 4 dwg



ФИГ.1

Область техники, к которой относится изобретение

Функции ранжирования, которые ранжируют документы в соответствии с их релевантностью к заданному поисковому запросу, известны. Продолжаются исследования в данной области техники по совершенствованию функций ранжирования, которые обеспечивают лучшие результаты поиска для заданного поискового запроса по сравнению с результатами поиска, сформированными поисковыми машинами, использующими известные функции ранжирования.

Сущность изобретения

Здесь описаны помимо всего прочего разнообразные технологии определения показателя релевантности документа для заданного документа в сети. Показатель релевантности документа формируется посредством функции ранжирования, которая содержит один или более независимых от запроса компонентов, при этом, по меньшей мере, один независимый от запроса компонент включает в себя параметр используемости, которым учитываются данные фактической используемости документа, содержащиеся и сохраняемые на web-сервере, для одного или более документов в сети. Функции ранжирования могут быть использованы поисковой машиной для ранжирования многочисленных документов по порядку (в типичном случае, в убывающем порядке), исходя из показателей релевантности этих многочисленных документов.

Данное краткое изложение сущности изобретения предоставлено, главным образом, для того, чтобы в упрощенной форме ознакомить читателя с одним или более предпочтительными концепциями, описанными ниже в разделе "Подробное описание". Это изложение сущности изобретения не предназначено для того, чтобы определять ключевые и/или необходимые признаки заявленного изобретения.

Перечень чертежей

Фиг.1 - иллюстративная логическая блок-схема, показывающая иллюстративные этапы в способе формирования ранжированных результатов поиска в ответ на введенный пользователем поисковый запрос.

Фиг.2 - блок-схема некоторых основных компонентов иллюстративного рабочего окружения для реализации раскрытых здесь способов и процессов.

Фиг.3 - логическая блок-схема, показывающая иллюстративные этапы в иллюстративном способе определения показателя релевантности документа для документов в сети.

Фиг.4 - логическая блок-схема, показывающая иллюстративные этапы в способе ранжирования результатов поиска, сформированных используя функцию ранжирования, содержащую параметр используемости.

Подробное описание изобретения

Чтобы ускорить понимание принципов способов и процессов, раскрытых в этом документе, следуют описания конкретных вариантов осуществления настоящего изобретения, и используется конкретный язык, чтобы описать конкретные варианты осуществления настоящего изобретения. Несмотря на это, должно быть понятно, что использование конкретной терминологии не означает ограничение объема раскрытых способов и процессов. Предусматривается, что изменения, дальнейшие модификации, а также дальнейшие применения принципов раскрытых способов и рассмотренных процессов очевидны обычным специалистам в данной области техники, к которой имеют отношение раскрытые способы и процессы.

Раскрываются способы определения показателя релевантности документа для документов в сети. Каждый показатель релевантности документа вычисляется,

используя функцию ранжирования, которая желательна содержит один или более независимых от запроса компонентов (например, компонент функции, который не зависит от данного поискового запроса или терма поискового запроса), один или более зависимых от запроса компонентов (например, компонент функции, который зависит от специфики, поискового запроса или терма поискового запроса), или их комбинацию. Определенные посредством функции ранжирования показатели релевантности документа могут быть использованы, чтобы ранжировать документы в пределах сетевого пространства (например, корпоративного пространства intranet) согласно каждому показателю релевантности документа. Иллюстративный процесс поиска, в котором могут быть использованы раскрытые способы, показан на Фиг.1 в виде иллюстративного процесса 10.

На Фиг.1 изображен иллюстративный процесс 10 поиска, который начинается с процесса на этапе 80, на котором пользователь вводит поисковый запрос. От этапа 80 иллюстративный процесс 10 поиска переходит к этапу 200, на котором поисковая машина осуществляет поиск всех документов в пределах сетевого пространства для одного или более термов поискового запроса. От этапа 200 иллюстративный процесс 10 поиска переходит к этапу 300, на котором функция ранжирования поисковой машины ранжирует документы в пределах сетевого пространства, исходя из показателя релевантности каждого документа, причем показатель релевантности документа основан на одном или более независимых от запроса компонентах, одном или более зависимых от запроса компонентах или их комбинации. От этапа 300 иллюстративный процесс 10 поиска переходит к этапу 400, на котором ранжированные результаты поиска представляются пользователю, в типичном случае в порядке убывания, идентифицируя документы в пределах сетевого пространства, которые наиболее релевантны поисковому запросу.

Как более подробно рассматривается ниже, в некоторых иллюстративных способах определения показателя релевантности документа, по меньшей мере, один независимый от запроса компонент функции ранжирования, используемый для определения показателя релевантности документа, принимает во внимание "данные используемости документа" или "статистические данные используемости документа", связанные с фактической используемостью одного или более документов в пределах сетевого пространства одним или более пользователями. Данные используемости документа и/или статистические данные формируются и сохраняются посредством прикладного кода на web-сервере, который является отдельным от заданной поисковой машины. Например, данные используемости документа могут обслуживаться web-сайтом так, чтобы каждый раз, когда пользователь запрашивает URL, сервер обновлял счетчик используемости. Счетчик используемости может обслуживать связанные с документом данные, полученные в течение заданного интервала времени, такого как прошлая неделя, прошлый месяц, прошлый год, либо время существования заданного документа или набора документов. Прикладной код может использоваться, чтобы получить с web-сайта данные используемости посредством (i) конкретного интерфейса прикладного программирования (API), (ii) запроса web-услуги или (iii) запроса web-страницы администрирования, которая возвращает данные используемости для каждого URL на web-сайте.

Для того чтобы формировать и обслуживать данные используемости в пределах сетевого пространства, а также хранить данные используемости в локальной или удаленной системе хранения, могут использоваться конкретные web-сайты. Подходящие для формирования, обслуживания и хранения данных используемости

документов web-сайты в пределах сетевого пространства включают в себя, но не ограничены ими, сайты символов WINDOWS® SHAREPOINT®.

Кроме того, раскрытые способы определения показателя релевантности документа могут использовать функцию ранжирования, которая содержит один или более дополнительных независимых от запроса компонентов. Подходящие дополнительные независимые от запроса компоненты включают в себя, но не в ограничительном смысле, независимый от запроса компонент, который принимает во внимание дистанцию в кликах (количество щелчков, например, мышью («кликов») от документа до страницы верхнего уровня) для каждого документа в пределах сетевого пространства, как описано в Американской Патентной Заявке № 10/955,983, озаглавленной "SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE" ("СИСТЕМА И СПОСОБ РАНЖИРОВАНИЯ РЕЗУЛЬТАТОВ ПОИСКА, ИСПОЛЬЗУЯ ДИСТАНЦИЮ В КЛИКАХ"), поданной 30 августа 2004 года, независимый от запроса компонент, который принимает во внимание смещенную дистанцию в кликах каждого документа в пределах сетевого пространства, как описано в Американской Патентной Заявке № 11/206,286, озаглавленной "RANKING FUNCTIONS USING A BIASED CLICK DISTANCE OF A DOCUMENT ON A NETWORK" ("ФУНКЦИИ РАНЖИРОВАНИЯ, ИСПОЛЬЗУЮЩИЕ СМЕЩЕННУЮ ДИСТАНЦИЮ В КЛИКАХ ДОКУМЕНТА В СЕТИ"), поданной 15 августа 2005 года, и независимый от запроса компонент, который принимает во внимание URL каждого документа в пределах сетевого пространства, как описано в Американской Патентной Заявке № 10/955,983, озаглавленной "SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE" ("СИСТЕМА И СПОСОБ РАНЖИРОВАНИЯ РЕЗУЛЬТАТОВ ПОИСКА, ИСПОЛЬЗУЯ ДИСТАНЦИЮ В КЛИКАХ"), поданной 30 августа 2004 года. Раскрытие вышеупомянутых Американских патентных заявок, права по которым переуступлены правообладателю по настоящей заявке, включено в настоящую патентную заявку посредством ссылки.

Кроме того, в дальнейшем иллюстративном варианте осуществления настоящего изобретения раскрытые способы определения показателя релевантности документа используют функцию ранжирования, которая содержит, по меньшей мере, один независимый от запроса компонент, который включает в себя и вышеописанный параметр используемости документа, и один или более из вышеописанных дополнительных независимых от запроса компонентов.

Показатель релевантности документа может быть использован для ранжирования документов в пределах сетевого пространства. Например, способ ранжирования документов в сети может содержать этапы, на которых: определяют показатель релевантности документа каждого документа в сети, используя вышеописанный способ; и ранжируют документы в желаемом порядке (в типичном случае, в порядке убывания), основываясь на показателях релевантности документа каждого документа.

Также, показатель релевантности документа может использоваться, чтобы ранжировать результаты поиска по поисковому запросу. Например, способ ранжирования результатов поиска по поисковому запросу может содержать этапы, на которых: определяют показатель релевантности документа для каждого документа в результатах поиска поискового запроса, используя вышеописанный способ, и ранжируют документы в желаемом порядке (в типичном случае, в порядке убывания), основываясь на показателях релевантности документа каждого документа.

Прикладные программы, использующие способы, раскрытые здесь, могут

загружаться и исполняться на разнообразии компьютерных систем, содержащих множество аппаратных компонентов. Иллюстративная компьютерная система и иллюстративная рабочая среда для осуществления на практике раскрытых здесь способов, описаны ниже.

Иллюстративная рабочая среда

Фиг.2 - иллюстрация примера подходящего окружения 100 вычислительной системы, в котором могут быть реализованы раскрытые здесь способы.

Окружение 100 вычислительной системы является только одним примером

подходящего вычислительного окружения и не предполагает какого-либо ограничения в отношении объема использования или функциональных возможностей раскрытых здесь способов. Также вычислительное окружение 100 не должно быть интерпретировано как подразумевающее какую-либо зависимость или требования, относящиеся к какому-либо компоненту или комбинации компонентов, проиллюстрированных в иллюстративном рабочем окружении 100.

Раскрытые здесь способы могут быть реализованы с многочисленными другими окружениями или конфигурациями вычислительных систем общего или специального назначения. Примеры известных вычислительных систем, окружений и/или конфигураций, которые могут быть применимы для использования с раскрытыми здесь способами, включают в себя, но не в ограничительном смысле, персональные компьютеры, компьютеры-серверы, карманные или портативные устройства, многопроцессорные системы, основанные на микропроцессоре системы, телевизионные приставки, программируемую бытовую электронику, сетевые персональные компьютеры (PC), миникомпьютеры, универсальные компьютеры, среды распределенных вычислений, которые включают в себя любые из вышеупомянутых систем или устройств, и т.п.

Раскрытые здесь способы и процессы могут быть описаны в общем контексте исполняемых компьютером инструкций, таких как исполняемые компьютером программные модули. Обычно, программные модули включают в себя процедуры, программы, объекты, компоненты, структуры данных и т.д., которые выполняют конкретные задачи или реализуют конкретные абстрактные типы данных. Раскрытые здесь способы и процессы также могут быть реализованы в средах распределенных вычислений, где задачи выполняются посредством удаленных устройств обработки, которые связаны посредством коммуникационной сети. В среде распределенных вычислений программные модули могут располагаться и на локальных, и на удаленных компьютерных носителях данных, включая запоминающие устройства памяти.

Согласно Фиг.2 иллюстративная система для реализации раскрытых здесь способов и процессов включает в себя универсальное вычислительное устройство в виде компьютера 110. Компоненты компьютера 110 могут включать в себя, но не в ограничительном смысле, блок 120 обработки, системную память 130 и системную шину 121, которая соединяет различные системные компоненты, включая, но не в ограничительном смысле, системную память 130, с блоком 120 обработки. Системная шина 121 может быть любой из нескольких типов шинных структур, включая шину памяти или контроллер памяти, периферийную шину и локальную шину, использующих любую из множества шинных архитектур. В качестве примера, а не ограничения, такие архитектуры включают в себя шину Архитектуры промышленного стандарта (ISA), шину Микроканальной архитектуры (MCA), шину Улучшенной ISA (EISA), локальную шину Видео Ассоциации Стандартов

Электроники (VESA) и локальную шину межсоединения Периферийных Устройств (PCI), так же известную, как шина Расширения.

В типичном случае компьютер 110 включает в себя множество машиночитаемых носителей. Машиночитаемые носители могут быть любыми доступными носителями информации, к которым можно обратиться посредством компьютера 110, и включают в себя как энергозависимые, так и энергонезависимые носители, как съемные, так и несъемные носители информации. В качестве примера, а не ограничения, машиночитаемые носители информации могут включать в себя компьютерные носители данных и коммуникационные среды. Компьютерные носители информации включают в себя энергозависимые и энергонезависимые носители информации, съемные и несъемные носители информации, реализованные любым способом или технологией хранения информации, такой как машиночитаемые инструкции, структуры данных, программные модули или другие данные. Компьютерные носители информации включают в себя, но не в ограничительном смысле, оперативную память (RAM), постоянную память (ROM), электронно-перепрограммируемую постоянную память (EEPROM), флэш-память или память другой технологии, компакт-диск, универсальный цифровой диск (DVD) или другой накопитель на оптических дисках, магнитные кассеты, магнитную ленту, накопитель на магнитных дисках или другие магнитные запоминающие устройства, либо любой другой носитель информации, который может использоваться для хранения желаемой информации и к которому можно обратиться посредством компьютера 110. В типичном случае коммуникационные среды воплощают машиночитаемые инструкции, структуры данных, программные модули или другие данные в модулированном информационном сигнале, таком как несущая волна или другой транспортный механизм, и включают в себя любые среды доставки информации. Термин "модулированный информационный сигнал", означает сигнал, одна или более характеристик которого устанавливаются или изменяются таким образом, чтобы кодировать информацию в этом сигнале. В качестве примера, а не ограничения, коммуникационные среды включают в себя проводные среды, такие как проводная сеть или прямое проводное соединенное, и беспроводные среды, такие как звуковые, радиочастотные, инфракрасные и другие беспроводные среды. Понятием «машиночитаемый носитель» как оно здесь используется, также охватываются комбинации любых из вышеупомянутых сред и носителей.

Системная память 130 включает в себя компьютерные носители информации в виде энергозависимой и/или энергонезависимой памяти, такой как постоянная память 131 (ROM) и оперативная память 132 (RAM). Базовая система 133 ввода-вывода (BIOS), содержащая базовые процедуры, которые помогают передавать информацию между элементами в пределах компьютера 110, например во время запуска, в типичном случае хранится в ROM 131. В типичном случае оперативная память 132 содержит данные и/или программные модули, которые непосредственно доступны и/или в настоящее время обрабатываются блоком 120 обработки. В качестве примера, а не ограничения Фиг.2 иллюстрирует операционную систему 134, прикладные программы 135, другие программные модули 136 и данные 137 программ.

Компьютер 110 также может включать в себя другие съемные/несъемные, энергозависимые/энергонезависимые компьютерные носители информации. Исключительно в качестве примера Фиг.2 иллюстрирует накопитель 140 на жестких дисках, который считывает или записывает на несъемные энергонезависимые магнитные носители, магнитный дисковод 151, который считывает или записывает на

съемный энергонезависимый магнитный диск 152, и оптический дисковод 155, который считывает или записывает на съемный, энергонезависимый оптический диск 156, такой как CD-ROM или другие оптические носители информации. Другие съемные/несъемные, энергозависимые/энергонезависимые компьютерные носители информации, которые могут использоваться в иллюстративном рабочем окружении, включают в себя, но не в ограничительном смысле, кассеты с магнитной лентой, карты флэш-памяти, универсальные цифровые диски, цифровую видеоленту, твердотельную RAM, твердотельную ROM и т.п. В типичном случае накопитель 141 на жестких дисках подключается к системной шине 121 посредством интерфейса несъемной памяти, такого как интерфейс 140, а магнитный дисковод 151 и оптический дисковод 155 в типичном случае подключаются к системной шине 121 посредством интерфейса съемной памяти, такого как интерфейс 150.

Накопители и дисководы, ассоциированные с ними компьютерные носители информации, рассмотренные выше и проиллюстрированные на Фиг.2, обеспечивают для компьютера 110 хранение машиночитаемых команд, структур данных, программных модулей и других данных. Например, на Фиг.2 накопитель 141 на жестких дисках иллюстрирован в качестве хранилища операционной системы 144, прикладных программ 145, других программных модулей 146 и данных 147 программ. Отметим, что эти компоненты могут быть либо такими же, либо отличающимися от операционной системы 134, прикладных программ 135, других программных модулей 136 и данных 137 программ. Операционной системе 144, прикладным программам 145, другим программным модулям 146 и данным 147 программ здесь присвоены другие ссылочные номера, чтобы проиллюстрировать, что они, как минимум, являются другими копиями.

Пользователь может ввести команды и информацию в компьютер 110 посредством устройств ввода, таких как клавиатура 162 и указательное устройство 161, обычно упоминаемое как мышь, шаровой манипулятор или сенсорная панель. Другие устройства ввода (не показаны) могут включать в себя микрофон, джойстик, игровую клавиатуру, спутниковую антенну, сканер и т.п. Эти и другие устройства ввода данных часто подключаются к блоку 120 обработки посредством интерфейса 160 пользовательского ввода, который соединен с системной шиной 121, но могут быть подключены посредством других интерфейсных и шинных структур, таких как параллельный порт, игровой порт или универсальная последовательная шина (USB). Монитор 191 или другой тип устройства отображения также подключен к системной шине 121 посредством интерфейса, такого как видеоинтерфейс 190. Кроме монитора 191, компьютер 110 также может включать в себя другие периферийные устройства вывода, такие как громкоговорители 197 и принтер 196, которые могут быть подключены посредством интерфейса 195 периферийного вывода.

Компьютер 110 может работать в сетевом окружении, используя логические соединения с одним или более удаленными компьютерами, такими как удаленный компьютер 180. Удаленный компьютер 180 может быть персональным компьютером, сервером, маршрутизатором, сетевым PC, одноранговым устройством или другим общим узлом сети и в типичном случае включает в себя многие или все из элементов, описанных выше в отношении компьютера 110, хотя на Фиг.2 проиллюстрировано только запоминающее устройство 181 памяти. Изображенные на Фиг.2 логические соединения включают в себя локальную сеть 171 (LAN) и глобальную сеть 173 (WAN), но также могут включать в себя другие сети. Такие сетевые окружения являются обычными в офисах, компьютерных сетях масштаба предприятия, сетях Интранет и

Интернет.

При использовании в сетевом окружении LAN, компьютер 110 подключается к LAN 171 посредством сетевого интерфейса или адаптера 170. При использовании в сетевом окружении WAN в типичном случае компьютер 110 включает в себя модем 172 или другие средства для установления связи по WAN 173, такой как Интернет. Модем 172, который может быть внутренним или внешним, может быть подключен к системной шине 121 посредством интерфейса 160 пользовательского ввода или другого подходящего механизма. В сетевом окружении программные модули, изображенные соответствующими компьютеру 110, или их части могут быть сохранены в удаленном запоминающем устройстве памяти. В качестве примера, а не ограничения, на Фиг.2 проиллюстрированы удаленные прикладные программы 185, находящиеся в устройстве 181 памяти. Должно быть понятно, что показанные сетевые соединения являются иллюстративными и могут быть использованы другие средства установления линии связи между компьютерами.

Раскрытые здесь способы и процессы могут быть реализованы, используя одну или более прикладных программ, включающих в себя, но не в ограничительном смысле, программное приложение серверной системы (например, программное приложение WINDOWS SERVER SYSTEM™), приложение ранжирования поиска и приложение для формирования, обслуживания и сохранения данных используемости документов в пределах сетевого пространства (например, приложение сервисов и WINDOWS® SHAREPOINT®), любое из которых может быть одной из многочисленных прикладных программ, обозначенных в иллюстративной системе 100, как прикладные программы 135, прикладные программы 145 и удаленные прикладные программы 185.

Как упомянуто выше, специалисты в данной области техники поймут, что раскрытые способы формирования показателя релевантности документа для заданного документа могут быть реализованы в других конфигурациях компьютерных систем, включая карманные устройства, многопроцессорные системы, основанную на микропроцессоре или программируемую бытовую электронику, сетевые персональные компьютеры, миникомпьютеры, универсальные компьютеры и т.п. Раскрытые способы формирования показателя релевантности документа для заданного документа также могут быть осуществлены на практике в средах распределенных вычислений, где задачи выполняются посредством удаленных устройств обработки, которые связаны посредством коммуникационной сети. В среде распределенных вычислений программные модули могут быть расположены и в локальных, и в удаленных запоминающих устройствах памяти.

Описание предпочтительных вариантов осуществления

Как рассмотрено выше, предоставлены способы определения показателя релевантности документа для документа в сети. Раскрытые способы могут ранжировать документ в сети, используя функцию ранжирования, которая принимает в расчет значение используемости документа каждого документа в сети.

Раскрытые способы определения показателя релевантности документа для документа в сети могут содержать некоторое количество этапов. В одном иллюстративном варианте осуществления настоящего изобретения способ определения показателя релевантности документа для документа в сети содержит этапы, на которых назначают фактическое значение (U_A) используемости одному или более документам в сети, содержащей N документов, при этом фактическое значение (U_A) используемости основано на обслуживаемых и хранящихся на сервере данных

фактической используемости; если меньше чем N документам назначено фактическое значение (U_A) используемости, документам, которые не имеют ассоциированных с ними данных фактической используемости, назначают используемое по умолчанию значение (U_D) используемости; и используют значение используемости каждого документа (то есть U_A или U_D), чтобы определить показатель релевантности документа заданного документа в сети.

Используемый здесь термин "данные фактической используемости" представляет собой один или более типов данных, ассоциированных с "используемостью" документа одним или более пользователями. Типы данных фактической используемости для данного документа или набора документов могут включать в себя, но не в ограничительном смысле, количество просмотров документа всеми пользователями в пределах заданного промежутка времени, среднее количество просмотров документа в расчете на пользователя в пределах заданного промежутка времени, полное время использования конкретного документа в пределах заданного промежутка времени, среднее время использования конкретного документа в пределах заданного промежутка времени и т.д. Упомянутый заданный промежуток времени может быть, например, прошлой неделей, прошлым месяцем, прошлым годом, временем существования документа либо любым другим желаемым периодом времени.

Этапы, на которых формируют, обслуживают и сохраняют данные используемости документа или статистические данные для документов в пределах сетевого пространства, могут быть выполнены посредством прикладного кода, обычно находящегося в вычислительных системах. Данные используемости документа формируются, обслуживаются и сохраняются независимо от заданного поискового запроса или поисковой машины и в типичном случае формируются, обслуживаются и сохраняются посредством прикладного кода на сервере, который обслуживает документ (или страницу) и делает документ (или страницу) доступным для пользователя. Прикладные программы, подходящие для формирования, обслуживания и сохранения данных используемости документа или статистических данных, включают в себя, но не в ограничительном смысле, сервисы WINDOWS® SHAREPOINT® и другие подобные прикладные программы.

К данным используемости документа, хранящимся и обслуживаемым на этих сайтах сервисов, так же как и на других web-сайтах, выполняющих подобную функцию, можно осуществить доступ, используя прикладной код, как рассмотрено выше. Например, к данным используемости документа можно обратиться с заданного web-сайта (например, сайта сервисов WINDOWS® SHAREPOINT®) посредством (i) специального интерфейса прикладного программирования (API), (ii) запроса web-сервиса или (iii) посредством запроса web-страницы администрирования, которая возвращает данные используемости каждого URL на web-сайте.

Раскрытые способы определения показателя релевантности документа для документа в сети могут содержать некоторое количество дополнительных этапов, включая, но не в ограничительном смысле, этапы, на которых осуществляют мониторинг одного или более документов в пределах сетевого пространства в отношении фактической используемости документа; сохраняют данные фактической используемости документа одного или более документов в локальном или удаленном файле хранения данных; вычисляют фактическое значение (U_A) используемости для документа на основании данных фактической используемости для документа или папки, содержащей документ; сохраняют фактические значения (U_A) используемости

для одного или более документов в локальном или удаленном файле хранения данных; запрашивают сохраненные данные используемости документа или фактические значения (U_A) используемости из локального или удаленного файла хранения данных (например, запрос таких данных поисковой машиной после конкретного поискового запроса пользователя); извлекают данные фактической используемости документа или фактическое значение (U_A) используемости для одного или более документов из локального или удаленного файла хранения данных; и в необязательном порядке объединяют значение используемости документа (то есть фактическое или используемое по умолчанию) с одним или более дополнительными свойствами документа для определения показателя релевантности документа для документа.

Фиг.3 представляет собой логическую блок-схему, показывающую иллюстративные этапы в иллюстративном способе предоставления фактических или используемых по умолчанию значений используемости для документов в сети, за которыми следует необязательная процедура понижения/повышения администратором системы. Как показано на Фиг.3, иллюстративный способ 401 начинается на этапе 402 и переходит к этапу 403. На этапе 403 первый документ в сети обрабатывается поисковым агентом для определения данных фактической используемости.

Этап обработки документа поисковым агентом для определения данных фактической используемости (этап 403) может быть выполнен, используя приложение поискового агента, выполненное с возможностью определять, имеет ли первый документ какие-нибудь ассоциированные с ним данные фактической используемости, и если первый документ имеет ассоциированные с ним данные фактической используемости, то извлекают эти данные фактической используемости. Приложения поискового агента, подходящие для использования в раскрытых способах предоставления фактических или используемых по умолчанию значений используемости документов в сети, включают в себя, но не в ограничительном смысле, приложения поискового агента, описанные в Американских Патентах № 6,463,455 и № 6,631,369, все содержимое которых полностью включено в настоящий документ посредством ссылки.

Как рассматривалось выше, данные фактической используемости могут быть получены из одного или более файлов, которые хранят данные фактической используемости одного или более документов в сети. Данные фактической используемости могут быть сохранены вместе с документом, как компонент документа или могут храниться в файле хранения данных, отдельном от фактического документа. Подходящие удаленные системы хранения включают в себя, но не в ограничительном смысле, серверы WINDOWS® SHAREPOINT® (WSS) коммерчески доступные продукты от Корпорации Microsoft (Редмонд, Вашингтон), а также любую другую подобную удаленную систему хранения. Например, удаленная система WSS хранения генерирует данные фактической используемости, включая, например, количество запросов к каждому документу в данной сети всеми пользователями, и вырабатывает статистику по количеству кликов по каждому документу в течение прошлой недели, прошлого месяца, прошлого года или всего времени существования документа либо любого другого промежутка времени. Кроме того, как отмечено выше, должно быть понятно, что раскрытые здесь способы не ограничены удаленной системой WSS хранения, и в раскрытых способах может использоваться удаленная система WSS хранения или любая другая подобная система данных документов.

Когда документ обработан поисковым агентом, иллюстративный способ 401

переходит к этапу 404 принятия решения. На этапе 404 принятия решения посредством прикладного кода определяют, имеет ли документ ассоциированные с ним данные фактической используемости. Если принято решение, что документ имеет ассоциированные с ним данные фактической используемости, иллюстративный способ 401 переходит к этапу 405, на котором документу назначается значение (U_A) используемости, основанное на фактической используемости. Фактическое значение (U_A) используемости может быть определено, используя один или более компонентов ассоциированных с документом данных фактической используемости. Например, в некоторых вариантах осуществления настоящего изобретения фактическое значение (U_A) используемости может быть связано только с количеством пользователей, просмотревших документ. В других вариантах осуществления настоящего изобретения присвоенное документу фактическое значение (U_A) используемости может быть связано с количеством просмотров документа всеми пользователями в пределах заданного промежутка времени, средним количеством просмотров документа в расчете на пользователя в пределах заданного промежутка времени, полным временем использования конкретного документа в пределах заданного промежутка времени, средним временем использования конкретного документа в пределах заданного промежутка времени либо комбинацией любых из вышеупомянутых критериев, причем упомянутый заданный промежуток времени включает в себя прошлую неделю, прошлый месяц, прошлый год, время существования документа или любой другой желаемый промежуток времени.

В некоторых случаях ассоциированные с заданным документом данные фактической используемости предполагают, что документ не использовался или не просматривался в течение заданного периода времени. В таком случае документу можно было бы назначить значение (U_A) используемости, равное нулю, которое указывает отсутствие используемости в течение этого периода времени, однако в типичном случае значениям (U_A) используемости, основанным на фактическом использовании или отсутствии фактического использования, присваивают число, отличное от нуля.

Кроме того, в некоторых случаях данные фактической используемости могут быть ассоциированы с набором документов в отличие от отдельных документов. Например, папка может содержать набор документов, и ассоциированный сервер может проследить только данные используемости, связанные с доступом (то есть использованием) к папке, а не отдельными документами в папке. В этом варианте осуществления настоящего изобретения, если есть ассоциированные с папкой данные фактической используемости, значение (U_A) используемости может быть предоставлено для каждого документа в папке, исходя из данных фактической используемости папки. В типичном случае каждое значение (U_A) используемости будет одинаковым для каждого документа в папке, однако, если это желательно, разным документам в папке могут быть назначены разные значения (U_A) используемости.

С этапа 405 иллюстративный способ 401 переходит к описанному ниже этапу 406 принятия решения.

Возвращаясь к этапу 404 принятия решения, если принято решение, что документ не имеет ассоциированных с ним данных фактической используемости, иллюстративный способ 401 переходит к этапу 407, на котором документу назначается используемое по умолчанию значение (U_D) используемости. Например, заданное по умолчанию значение (U_D) используемости может быть назначено документу, который является частью web-сайта, который не поддерживает данные используемости документа.

Назначенное документу используемое по умолчанию значение (U_D) используемости может использоваться, чтобы придать начальную значимость документу относительно документов, имеющих данные фактической используемости. Например, если более высокое значение используемости для заданного документа указывает относительную значимость этого документа в пределах сети, назначение документу более низкого используемого по умолчанию значения (U_D) используемости понижает значимость этого документа относительно других документов в сети.

В одном иллюстративном варианте осуществления настоящего изобретения, в котором более высокое значение используемости данного документа указывает относительную значимость этого документа в пределах сети, используемое по умолчанию значение (U_D) используемости может быть назначено документу относительно фактических значений (U_A) используемости, назначенных другим документам в сети. Например, чтобы понизить относительную значимость документа, документу может быть назначено используемое по умолчанию значение (U_D) используемости, причем это используемое по умолчанию значение (U_D) используемости меньше, чем любое фактическое значение (U_A) используемости, назначенное другим документам в сети, как описано выше. Если желательно увеличить относительную значимость документа, документу может быть назначено используемое по умолчанию значение (U_D) используемости больше, чем любое фактическое значение (U_A) используемости, назначенное другим документам в сети, или больше, чем некоторые из фактических значений используемости, назначенных нескольким другим документам в сети.

В других вариантах осуществления настоящего изобретения используемое по умолчанию значение (U_D) используемости может быть назначено документу без данных фактической используемости так, чтобы документу задавалась средняя относительная значимость по сравнению с документами, имеющими назначенное фактическое значение (U_A) используемости. Например, в этом варианте осуществления настоящего изобретения используемые по умолчанию значения (U_D) используемости документов без данных фактической используемости могут варьироваться от минимального назначенного фактического значения (U_{Amin}) используемости до максимального назначенного фактического значения (U_{Amax}) используемости или быть в пределах конкретного диапазона между минимальным назначенным фактическим значением (U_{Amin}) используемости и максимальным назначенным фактическим значением (U_{Amax}) используемости. В этом варианте осуществления настоящего изобретения документом без данных фактической используемости надеется средняя относительная значимость, предлагая среднюю используемость, по сравнению с документами, которые имеют ассоциированные с ними данные фактической используемости.

От этапа 407 иллюстративный способ 401 переходит к этапу 406 принятия решения. На этапе 406 принятия решения с помощью прикладного кода делается определение того, все ли документы в сети имеют фактическое (U_A) или используемое по умолчанию значение (U_D) используемости. Если принято решение, что не все документы в сети имеют фактическое (U_A) или используемое по умолчанию значение (U_D) используемости, иллюстративный способ 401 переходит к этапу 408, на котором следующий документ обрабатывается поисковым агентом для определения данных фактической используемости. От этапа 408 иллюстративный способ 401 возвращается

к этапу 404 принятия решения и действует, как рассмотрено выше.

Возвращаясь к этапу 406 принятия решения, если с помощью прикладного кода определено, что все документы в сети имеют фактическое (U_A) или используемое по умолчанию значение (U_D) используемости, иллюстративный способ 401 переходит к
 5 этапу 409 принятия решения. На этапе 409 системным администратором принимается решение, понизить ли какие-нибудь фактические (U_A) или используемые по умолчанию значения (U_D) используемости, чтобы более точно отразить значимость заданного документа в пределах сетевого пространства. Если принято решение понизить одно
 10 или более фактических (U_A) или используемых по умолчанию значений (U_D) используемости, чтобы более точно отобразить значимость одного или более документов в пределах сетевого пространства, иллюстративный способ 401 переходит к этапу 410, на котором фактическое (U_A) или используемое по умолчанию значение
 15 (U_D) используемости одного или более документов (или URL) корректируют либо с уменьшением, либо с увеличением. От этапа 410 иллюстративный способ 401 переходит к описанному ниже этапу 411.

Возвращаясь к этапу 409 принятия решения, если принято решение не понижать (или не повышать) одно или более фактических (U_A) или используемых по умолчанию
 20 (U_D) значений используемости, иллюстративный способ 401 переходит к этапу 411. На этапе 411 фактическое (U_A) и используемое по умолчанию (U_D) значения используемости используются в функции ранжирования для того, чтобы определить общий показатель релевантности документа для каждого документа в пределах
 25 сетевого пространства. От этапа 411 иллюстративный способ 401 переходит к этапу 412 конец.

Когда все фактические (U_A) и используемые по умолчанию (U_D) значения используемости определены и в необязательном порядке понижены (или в необязательном порядке повышены), если это желательно, фактическое (U_A) или
 30 используемое по умолчанию (U_D) значение используемости каждого документа может использоваться как параметр в функции ранжирования, чтобы предоставить показатель релевантности документа для каждого документа. Такой показатель релевантности документа может использоваться, чтобы ранжировать результаты
 35 поиска по поисковому запросу. На Фиг.4 показан иллюстративный способ ранжирования результатов поиска, сформированных, используя функцию ранжирования, содержащую параметр в виде значения используемости документа.

Фиг.4 представляет собой логическую блок-схему, показывающую иллюстративные этапы в иллюстративном способе 20, при этом иллюстративный способ 20 содержит в
 40 себе способ ранжирования результатов поиска, сформированных с использованием функции ранжирования, содержащей параметр в виде значения используемости. Как показано на Фиг.4, иллюстративный способ 20 начинается на этапе 201 и переходит к этапу 202. На этапе 202 посредством ввода поискового запроса пользователь
 45 запрашивает поиск. До этапа 202 предварительно были вычислены фактические или используемые по умолчанию значения используемости каждого документа в сети. От этапа 202 иллюстративный способ 20 переходит к этапу 203.

На этапе 203 фактическое или используемое по умолчанию значение используемости для каждого документа в сети объединяется с любыми другими
 50 статистическими данными документа (например, другими независимыми от запроса статистическими данными) для каждого документа, сохраненного в индексе. Объединение фактических или используемых по умолчанию значений используемости с другими статистическими данными документа дает возможность уменьшить время

отклика на запрос, так как вся информация, связанная с ранжированием, сгруппирована. Соответственно, после объединения каждый документ, перечисленный в индексе, имеет ассоциированное фактическое или используемое по умолчанию значение используемости. Как только объединение завершено, иллюстративный способ 20 переходит к этапу 204.

На этапе 204 независимые от запроса статистические данные документа для заданного документа, включая параметр используемости, предоставляются в виде компонента функции ранжирования. Зависимые от запроса данные также предоставляются для данного документа, в типичном случае в виде отдельного компонента функции ранжирования. Зависимые от запроса данные или связанная с содержимым часть функции ранжирования зависит от фактических условий поиска и содержимого данного документа.

В одном варианте осуществления настоящего изобретения функция ранжирования включает в себя, по меньшей мере, один независимый от запроса компонент (QID), содержащий параметр используемости. В одном варианте осуществления настоящего изобретения независимый от запроса компонент (QID) может быть представлен посредством следующего уравнения:

$$QID(doc) = w_u \frac{k_u U}{k_u + U} \quad (1)$$

в котором:

U представляет собой фактическое значение используемости или используемое по умолчанию значение используемости и w_u и k_u представляют собой параметры, корректирующие значение используемости. В дальнейшем варианте осуществления настоящего изобретения независимый от запроса компонент (QID) может быть представлен посредством следующего уравнения:

$$QID(doc) = w_u U + k_u \quad (2)$$

в котором:

U представляет собой фактическое значение используемости или используемое по умолчанию значение используемости и w_u и k_u представляют собой параметры, корректирующие значение используемости. В еще одном дальнейшем варианте осуществления настоящего изобретения независимый от запроса компонент (QID) может быть представлен посредством следующего уравнения:

$$QID(doc) = w_u [1 + \exp(-k_u U - B)] + C \quad (3)$$

в котором:

U представляет собой фактическое значение используемости или используемое по умолчанию значение используемости и w_u , k_u , B и C представляют собой параметры, корректирующие значение используемости (то есть скалярные константы).

В дальнейшем варианте осуществления настоящего изобретения функция ранжирования содержит сумму вышеописанных, независимого от запроса компонента (QID) и, по меньшей мере, одного зависимого от запроса

компонента (QD), такую как

$$Score = QD(doc, query) + QID(doc)$$

Компонент QD может быть любой функцией оценки документа. В одном варианте осуществления настоящего изобретения компонент QD соответствует взвешенной по полям функции оценки, описанной в американской патентной заявке № 10/804,326, поданной 18 марта 2004 года, озаглавленной "FIELD WELGHTING IN TEXT DOCUMENT SEARCHING" ("ВЗВЕШИВАНИЕ ПО ПОЛЯМ ПРИ ПОИСКЕ ТЕКСТОВЫХ ДОКУМЕНТОВ"), все содержимое которой включено в настоящий документ посредством ссылки. В соответствии с Американской патентной заявкой № 10/804,326 одним уравнением, которое может быть использовано в качестве задания взвешенной по полям функции оценки, является следующее:

$$QID(doc, query) = \sum \frac{wtf'(k_1 + 1)}{k_1 + wtf'} \times \log\left(\frac{N}{n}\right)$$

в котором:

wtf' представляет собой повторяемость взвешенного термина или сумму повторяемостей заданных терминов в поисковом запросе, умноженную на вес по всем полям (например, заглавие, основная часть и т.п. документа) и нормированную в соответствии с длиной каждого поля и соответствующей средней длиной,

N представляет собой количество документов в сети,

n представляет собой количество документов, содержащих терм запроса, и

k_1 - настраиваемая константа. Кроме того, вышеупомянутые термины и уравнения подробно описаны в американской патентной заявке № 10/804,326, все содержимое которой включено в настоящий документ посредством ссылки.

В некоторых вариантах осуществления настоящего изобретения функция ранжирования дополнительно может включать в себя компонент QID, который принимает в расчет (i) значение дистанции в кликах, которое определяется посредством способов, раскрытых в американской патентной заявке № 10/955,983, озаглавленной "SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE" ("СИСТЕМА И СПОСОБ РАНЖИРОВАНИЯ РЕЗУЛЬТАТОВ ПОИСКА, ИСПОЛЬЗУЯ ДИСТАНЦИЮ В КЛИКАХ"), поданной 30 августа 2004 года, (ii) значение смещенной дистанции в кликах, которое определяется посредством способов, раскрытых в американской патентной заявке № 11/206,286, озаглавленной "RANKING FUNCTIONS USING A BIASED CLICK DISTANCE OF A DOCUMENT ON A NETWORK" ("ФУНКЦИИ РАНЖИРОВАНИЯ, ИСПОЛЬЗУЮЩИЕ СМЕЩЕННУЮ ДИСТАНЦИЮ В КЛИКАХ ДОКУМЕНТА В СЕТИ"), поданной 15 августа 2005 года, все содержимое которых включено в этот документ посредством ссылки, (iii) глубину URL документа или (iv) значение, являющееся комбинацией (i) или (ii) и (iii). Например, этот необязательный дополнительный компонент QID может представлять собой функцию, как показано в уравнении:

$$QID(doc) = w_{cd} \frac{k_{od} \frac{CD}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD}}{k_{cd} + \frac{b_{cd} + b_{ud}}{k_{ew}}}$$

в котором:

w_{cd} представляет собой вес независимого от запроса компонента, такого как компонент, содержащий параметр дистанции в кликах или смещенной дистанции в

кликах,

b_{cd} представляет собой вес дистанции в кликах или смещенной дистанции в кликах относительно глубины URL,

b_{ud} представляет собой вес глубины URL,

CD представляет собой вычисленную или назначенную дистанцию в кликах или смещенную дистанцию в кликах для документа,

k_{ew} представляет собой корректирующую константу, которая определяется посредством оптимизации точности функции ранжирования, подобно другим корректирующим параметрам (то есть k_{ew} может представлять собой значение веса ребра, когда все ребра имеют одинаковое значение веса ребра либо k_{ew} может представлять собой обычное или среднее значение ребра, когда значения веса ребра отличаются друг от друга),

UD представляет собой глубину URL и

k_{cd} - константа насыщенности дистанции в кликах.

Взвешенные термы (w_{cd} , b_{cd} и k_{cd}) помогают определению, значимости каждого из относящихся к ним термов (то есть компонентов, содержащих параметр дистанции в кликах или смещенной дистанции в кликах, значение дистанции в кликах или смещенной дистанции в кликах для заданного документа и глубину URL данного документа, соответственно) и, разумеется, результата функции оценки.

Глубина URL (UD) является необязательным добавлением к вышеупомянутому независимому от запроса компоненту для сглаживания влияния, которое значение дистанции в кликах или смещенной дистанции в кликах могут оказывать на функцию оценки. Например, в некоторых случаях документ, который не очень значим (то есть имеет большую глубину URL), может иметь короткую дистанцию в кликах или смещенную дистанцию в кликах. Глубина URL представляется некоторым количеством косых черт в URL документа. Например, www.example.com/d1/d2/d3/d4.htm включает в себя четыре косых черты и поэтому URL имел бы глубину 4. Однако этот документ может иметь ссылку непосредственно с главной страницы www.example.com, предоставляя, тем самым, относительно низкое значение дистанции в кликах или смещенной дистанции в кликах. Включение терма глубины URL в вышеупомянутую функцию и взвешивание терма глубины URL по отношению к значению дистанции в кликах или смещенной дистанции в кликах компенсирует относительно высокое значение дистанции в кликах или смещенной дистанции в кликах, чтобы более точно отразить значимость документа в пределах сети. В зависимости от сети глубину URL, равную 3 или больше, можно считать глубокой ссылкой.

В одном варианте осуществления настоящего изобретения функция ранжирования, используемая для определения показателя релевантности документа для заданного документа, представляет собой следующую функцию:

$$Score = \sum \frac{wtf'(k_1 + 1)}{k_1 + wtf'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U}$$

термы которой описаны выше.

В других вариантах осуществления настоящего изобретения из функции ранжирования может быть удалена глубина URL или к функции ранжирования могут быть добавлены другие компоненты, чтобы увеличить точность зависимого от запроса компонента, независимого от запроса компонента или их обоих. Кроме того, вышеописанный независимый от запроса компонент, содержащий параметр

используемости, может быть включен в другие функции ранжирования (не показаны), чтобы улучшить ранжирование результатов поиска.

Как только на этапе 204 статистические данные документа для заданного документа предоставлены функции ранжирования, иллюстративный способ 20 переходит к этапу 205. На этапе 205 для заданного документа показатель релевантности документа определяют, сохраняют в памяти и ассоциируют с данным документом. От этапа 205 иллюстративный способ 20 переходит к этапу 206 принятия решения.

На этапе 206 принятия решения посредством прикладного кода определяют, был ли вычислен показатель релевантности документа для каждого документа в пределах сети. Если определено, что показатель релевантности документа не был вычислен для каждого документа в пределах сети, иллюстративный способ 20 возвращается к этапу 204 и продолжает действовать, как описано выше. Если определено, что показатель релевантности документа был вычислен для каждого документа в пределах сети, иллюстративный способ 20 переходит к этапу 207.

На этапе 207 результаты поиска по запросу, содержащие многочисленные документы, ранжируются согласно ассоциированным с ними показателями релевантности документа. Результирующие показатели релевантности документа принимают в расчет фактическое или используемое по умолчанию значение используемости каждого из документов в пределах сети. Как только результаты поиска ранжированы, иллюстративный способ 20 переходит к этапу 208, на котором ранжированные результаты отображают пользователю. От этапа 208 иллюстративный способ 20 переходит к этапу 209, на котором пользователем выбираются и просматриваются результаты с наивысшим рангом. От этапа 209 иллюстративный способ 20 переходит к этапу 210, на котором иллюстративный способ 20 заканчивается.

В дополнение к вышеописанным способам формирования показателя релевантности документа для документов в пределах сети и использования показателей релевантности документа, чтобы ранжировать результаты поиска по поисковому запросу, машиночитаемые носители, имеющие сохраненные в них исполняемые компьютером инструкции для выполнения вышеописанных способов, также раскрыты в этом документе.

Здесь также раскрыты вычислительные системы. Иллюстративная вычислительная система содержит, по меньшей мере, один модуль приложения, используемый в вычислительной системе, причем этот, по меньшей мере, один модуль приложения включает в себя загружаемый им прикладной код, который выполняет способ формирования показателя релевантности документа для документов в пределах сети. Прикладной код может быть загружен в вычислительную систему, используя любой из вышеописанных машиночитаемых носителей, содержащий исполняемые компьютером инструкции для формирования показателя релевантности документа для документов в пределах сети и использования показателей релевантности документа для ранжирования результатов поиска поискового запроса, как описано выше.

Несмотря на то, что описание было подробно изложено относительно конкретных вариантов осуществления настоящего изобретения, необходимо принять во внимание, что специалисты в данной области техники после достижения понимания вышеупомянутого могут легко представить себе преобразования, разновидности и эквиваленты этих вариантов осуществления настоящего изобретения. Следовательно,

объем раскрытых способов, машиночитаемых носителей и вычислительных систем должен быть определен прилагаемой формулой изобретения и любыми ее эквивалентами.

5

Формула изобретения

1. Машиночитаемый носитель, имеющий сохраненные на нем исполняемые компьютером инструкции, содержащие

инструкции для ранжирования документов в сети, при этом упомянутые
 10 исполняемые компьютером инструкции используют функцию ранжирования, используемую при запросе пользователем документов, которая содержит один или более независимых от запроса компонентов, причем, по меньшей мере, один независимый от запроса компонент включает в себя параметр используемости,
 15 которым учитываются сформированные сервером и сохраненные на сервере данные используемости для одного или более документов в сети, причем данные используемости сформированы и сохранены сетевой системой хранения, управляющей и обеспечивающей пользователям сетевой доступ к документам, при этом данные используемости содержат измерения фактического пользовательского взаимодействия
 20 через сетевую систему хранения с одним или более документами в сети так, что данные используемости документа отражают независимую от запроса интерактивную используемость документа множеством пользователей через сетевую систему хранения, причем исполняемые компьютером инструкции также назначают
 25 показатель, сформированный функцией ранжирования, для ранжирования документов в сети, при этом упомянутый показатель используют для ранжирования документов по порядку, при этом исполняемые компьютером инструкции дополнительно содержат инструкции для приема поискового запроса, содержащего поисковую строку, введенную пользователем; проведения поиска документов в сети для
 30 формирования результатов поиска, содержащих несколько документов; ранжирования нескольких документов из результатов поиска с использованием функции ранжирования для формирования ранжированных результатов поиска так, что документы ранжируют согласно и их данным относительной используемости, и их релевантности относительно поисковой строки; и обеспечения ранжированных
 35 результатов поиска пользователю.

2. Машиночитаемый носитель по п.1, в котором значение используемости включает в себя (i) фактическое значение используемости, основанное на данных фактической используемости, поддерживаемых сервером, или (ii) используемое по умолчанию
 40 значение используемости, которое не основано на данных фактической используемости.

3. Машиночитаемый носитель по п.1, в котором упомянутый, по меньшей мере, один независимый от запроса компонент представлен формулой:

$$45 \quad QID(doc) = w_u \frac{k_u U}{k_u + U},$$

в которой U - фактическое значение используемости или используемое по умолчанию значение используемости; и

w_u и k_u - параметры, корректирующие значение используемости.

50

4. Машиночитаемый носитель по п.1, в котором упомянутый, по меньшей мере, один независимый от запроса компонент включает в себя и (i) параметр используемости, и (ii) параметр дистанции в кликах (количества щелчков, например, мышью («кликов») от документа до страницы верхнего уровня) или смещенной

дистанции в кликах.

5. Машиночитаемый носитель по п.1, в котором упомянутый, по меньшей мере, один независимый от запроса компонент включает в себя и параметр используемости, и параметр глубины URL.

6. Машиночитаемый носитель по п.1, в котором показатель каждого документа формируется, используя формулу:

$$\text{Score} = \sum \frac{\text{wtf}(k_1 + 1)}{k_1 + \text{wtf}} \cdot \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U},$$

в которой wtf - взвешенная повторяемость термина;

N - количество документов в сети;

n - количество документов, содержащих поисковый терм;

w_{cd} - вес независимого от запроса компонента;

b_{cd} - вес дистанции в кликах;

b_{ud} - вес глубины URL;

CD - вычисленная дистанция в кликах или назначенная смещенная дистанция в кликах для документа;

k_{ew} - корректирующая константа, соответствующая весам ребер;

UD - глубина URL;

U - фактическое значение используемости или используемое по умолчанию значение используемости;

w_u и k_u - параметры, корректирующие значение используемости, и

k_{cd} и k_l - константы.

7. Машиночитаемый носитель по п.1, дополнительно содержащий исполняемые компьютером инструкции для предоставления возможности администратору вручную настраивать результаты ранжирования, сформированные посредством функции ранжирования.

8. Вычислительная система, содержащая, по меньшей мере, один модуль приложения, используемый в вычислительной системе, причем этот, по меньшей мере, один модуль приложения содержит прикладной код, загружаемый в вычислительную систему с машиночитаемого носителя по п.1.

9. Машиночитаемый носитель, имеющий сохраненные на нем исполняемые компьютером инструкции, содержащие

инструкции для ранжирования документов в сети, при этом упомянутые исполняемые компьютером инструкции используют функцию ранжирования, которая содержит один или более независимых от запроса компонентов, причем, по меньшей мере, один независимый от запроса компонент включает в себя параметр используемости, которым учитываются сформированные сервером и сохраненные на сервере данные используемости для одного или более документов в сети, при этом данные используемости содержат измерения фактического пользовательского взаимодействия с одним или более документами в сети, причем исполняемые компьютером инструкции также назначают показатель, сформированный функцией ранжирования, для ранжирования документов в сети, при этом упомянутый показатель используют для ранжирования документов по порядку, причем данные используемости содержат значения используемости соответствующих документов, причем значение используемости документа зависит от одного или более связанных с

используемостью свойств документа или папки, содержащей набор документов, причем упомянутые одно или более связанных с используемостью свойств включают в себя общее количество просмотров пользователями документа или папки в пределах заданного промежутка времени, среднее количество просмотров документа или папки в расчете на пользователя в пределах заданного промежутка времени, полное время использования конкретного документа или папки в пределах заданного промежутка времени, среднее время использования конкретного документа или папки в пределах заданного промежутка времени, при этом упомянутый заданный промежуток времени включает в себя прошлую неделю, прошлый месяц, прошлый год, время существования документа или папки, либо любой другой промежуток времени.

10. Машиночитаемый носитель по п.9, в котором одно или более связанных с используемостью свойств включают в себя общее количество просмотров пользователями документа или папки в пределах заданного промежутка времени.

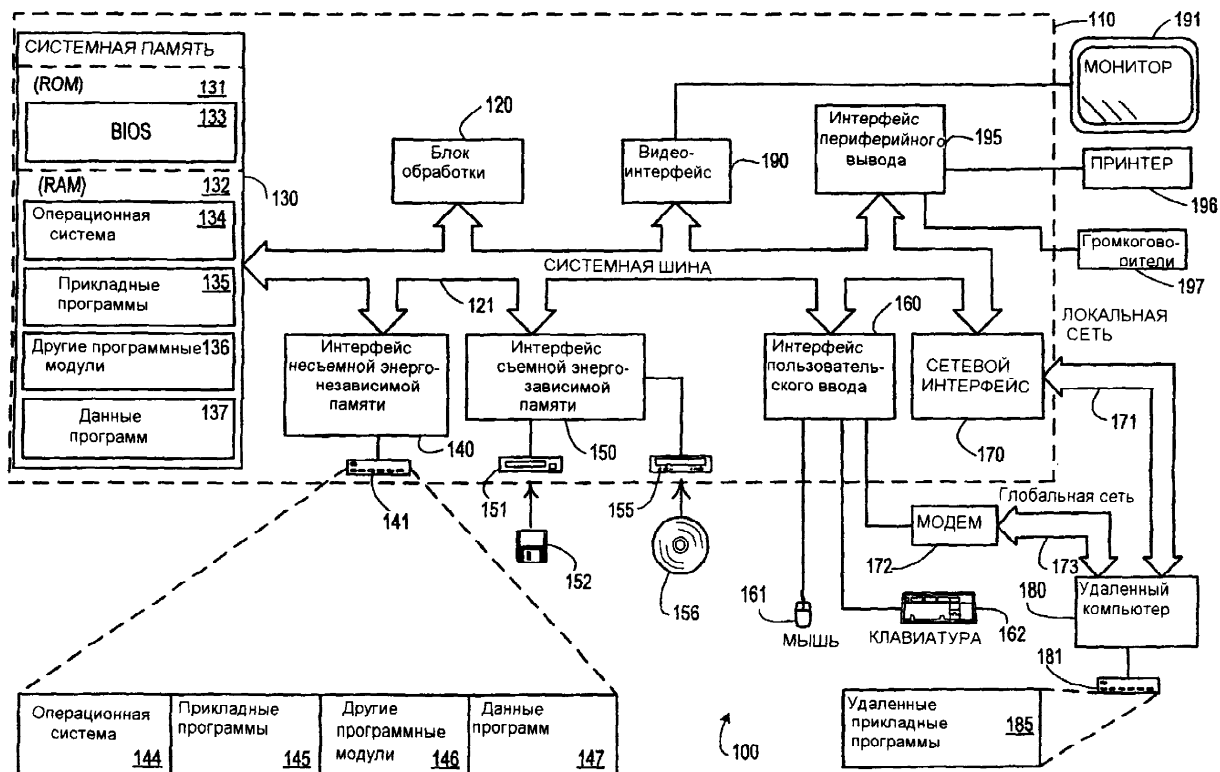
11. Машиночитаемый носитель по п.9, в котором одно или более связанных с используемостью свойств включают в себя среднее количество просмотров документа или папки в расчете на пользователя в пределах заданного промежутка времени.

12. Машиночитаемый носитель по п.9, в котором одно или более связанных с используемостью свойств включают в себя полное время использования конкретного документа или папки в пределах заданного промежутка времени.

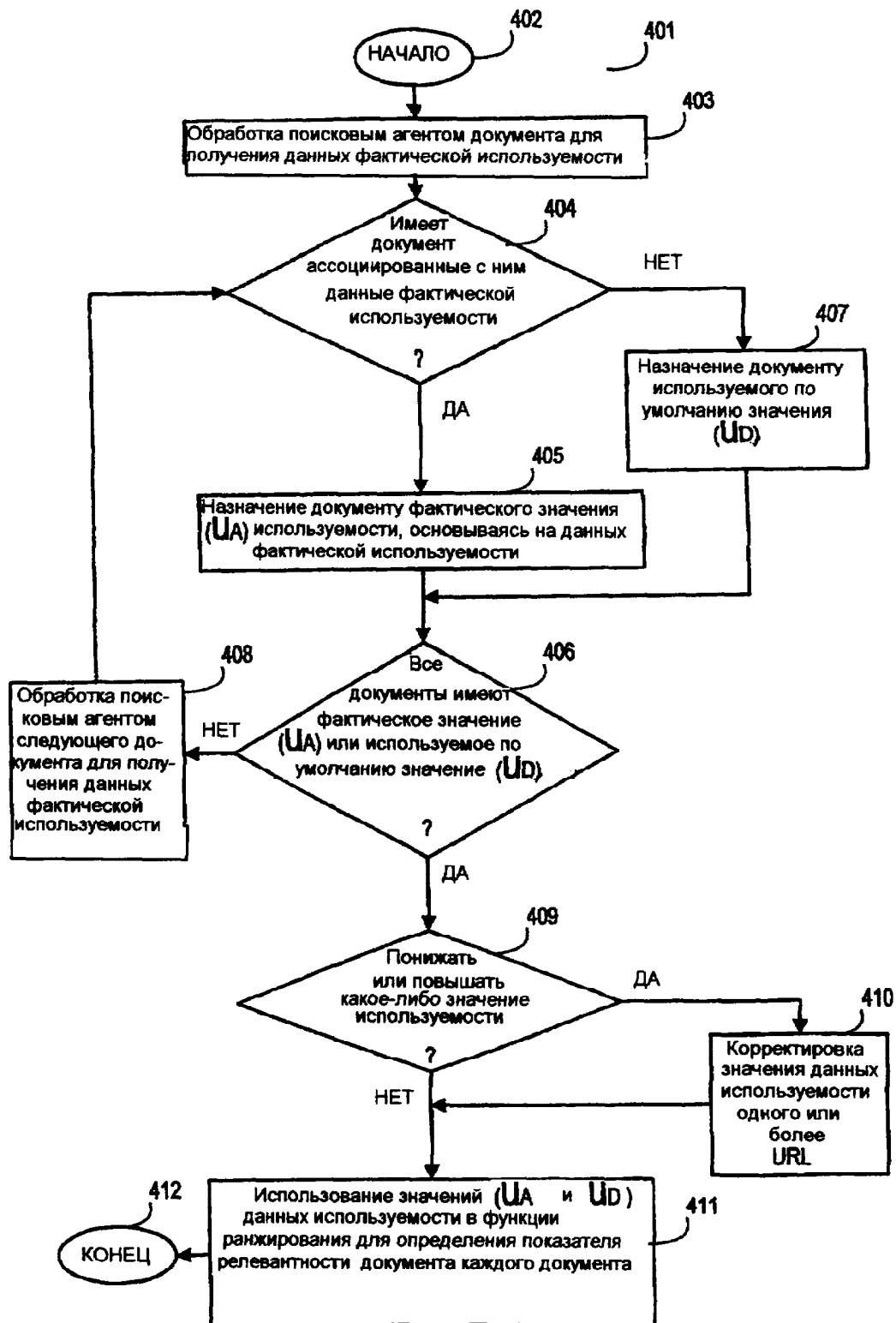
13. Машиночитаемый носитель по п.9, в котором одно или более связанных с используемостью свойств включают в себя среднее время использования конкретного документа или папки в пределах заданного промежутка времени.

14. Компьютерно-реализуемый способ ранжирования документов в сети, содержащий этапы, на которых:

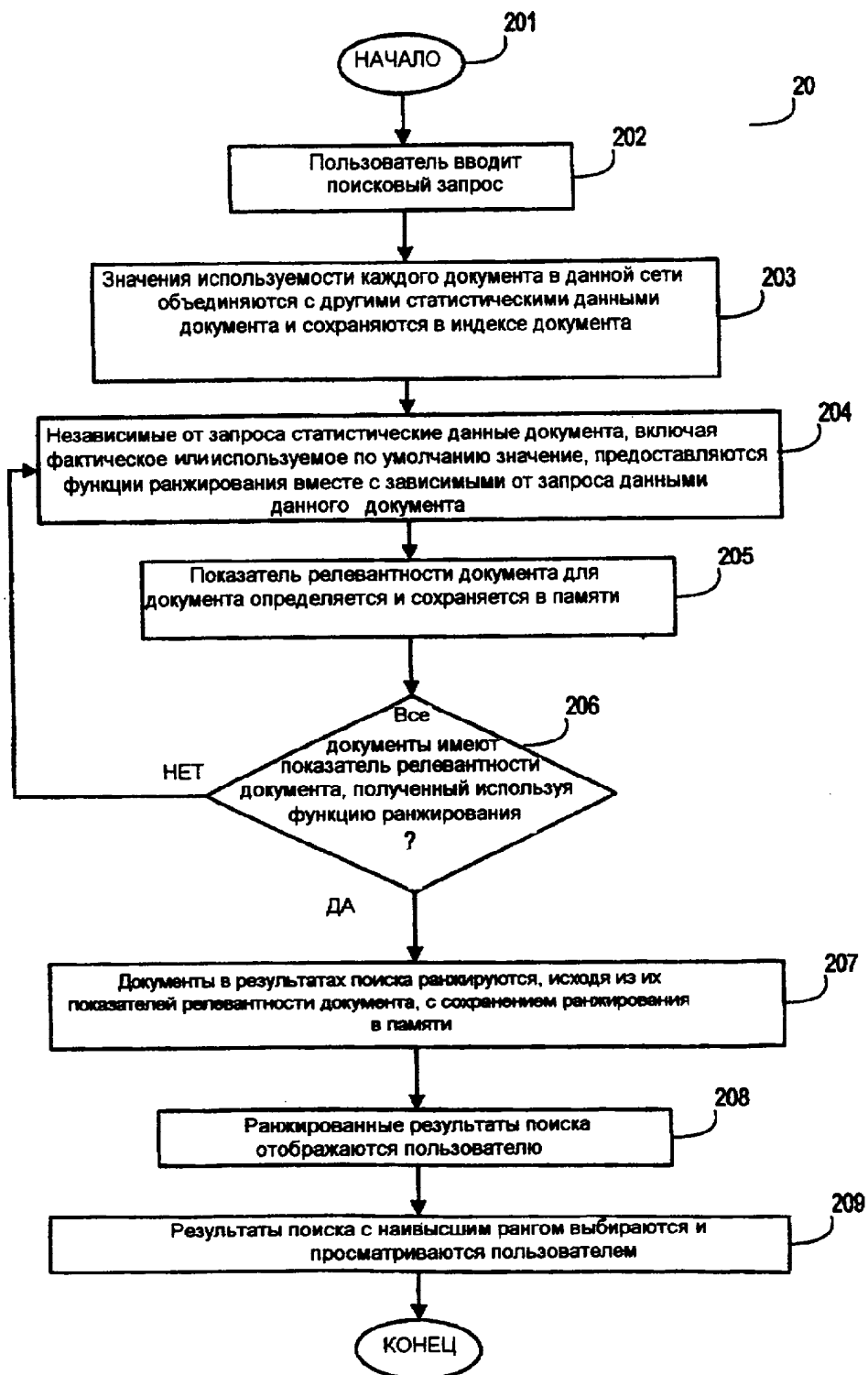
обеспечивают функцию ранжирования, которая содержит один или более независимых от запроса компонентов, причем, по меньшей мере, один независимый компонент включает в себя параметр используемости, которым учитываются сформированные сервером и сохраненные на сервере данные используемости для одного или более документов в сети, причем данные используемости содержат измерения фактического взаимодействия пользователя с одним или более документами в сети, и назначают показатель, сформированный функцией ранжирования, для ранжирования документов в сети, при этом упомянутый показатель используют для ранжирования документов по порядку, причем значение используемости зависит от одного или более связанных с используемостью свойств документа или папки, содержащей набор документов, причем упомянутые одно или более связанных с используемостью свойств включают в себя общее количество просмотров пользователями документа или папки в пределах заданного промежутка времени, среднее количество просмотров документа или папки в расчете на пользователя в пределах заданного промежутка времени, полное время использования конкретного документа или папки в пределах заданного промежутка времени, среднее время использования конкретного документа или папки в пределах заданного промежутка времени, при этом упомянутый заданный промежуток времени включает в себя прошлую неделю, прошлый месяц, прошлый год, время существования документа или папки, либо любой другой промежуток времени.



ФИГ.2



ФИГ.3



ФИГ. 4