



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G06F 16/338 (2019.08); G06F 16/31 (2019.08); G06F 16/93 (2019.08)

(21)(22) Заявка: 2019100812, 10.01.2019

(24) Дата начала отсчета срока действия патента:
10.01.2019

Дата регистрации:
05.08.2020

Приоритет(ы):

(22) Дата подачи заявки: 10.01.2019

(43) Дата публикации заявки: 10.07.2020 Бюл. № 19

(45) Опубликовано: 05.08.2020 Бюл. № 22

Адрес для переписки:

660077, г. Красноярск, ул. Батурина, 30/3-80,
ООО "Максиол", директору

(72) Автор(ы):

Багаев Максим Сергеевич (RU),
Багаева Ольга Львовна (RU)

(73) Патентообладатель(и):

Общество с ограниченной ответственностью
"МАКСИОЛ" (RU)

(56) Список документов, цитированных в отчете
о поиске: RU 2388050 C2, 27.04.2010. GB
2338324 A, 15.12.1999. US 6738678 B1, 18.05.2004.
US 2009/0327271 A1, 31.12.2009. US 2011/0265189
A1, 27.10.2011.

(54) Способ извлечения информации и корпоративная система поиска информации

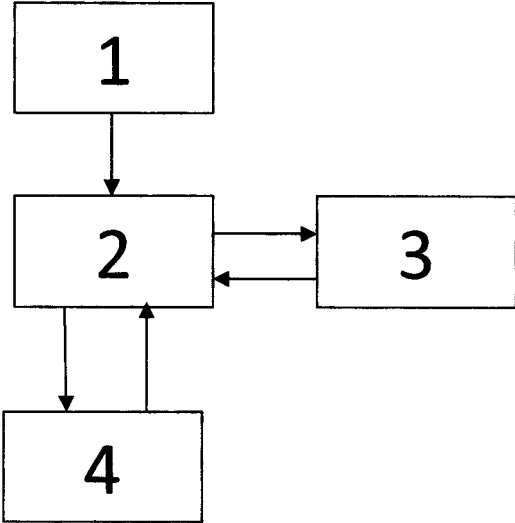
(57) Реферат:

Изобретение относится к области вычислительной техники. Техническим результатом является предоставление возможности поиска информации с обеспечением агрегирования информации, формы представления информации и соответствия требованиям информационной безопасности (ИБ) при предоставлении доступа к документам. Раскрыт способ извлечения информации, реализуемый в корпоративной системе поиска информации, выполненной с возможностью извлечения результатов из совокупности различных источников, содержащий этапы, на которых на входе принимают запрос; на выходе выдают этот запрос в совокупности различных источников, при этом он характеризуется иерархическим мульти-древовидным списочным представлением результатов поиска в виде одного или многих наборов связанных графов, узлы которых не содержат циклы, но могут иметь дубли и пересечения с другими наборами или ветвями, при этом могут иметь неуникальный

путь, характеризующийся последовательностью узлов, от корня до узла, корни деревьев имеют списочное представление и могут раскрываться при необходимости или при взаимодействии с пользователем интерактивно; включает выполнение поисковых запросов; проведение разбора поискового запроса при помощи морфологического анализа; обращение к базе поискового индекса через корпоративную систему поиска информации и извлечение из нее результатов поиска; формирование результатов поиска, используя корпоративную систему поиска информации, исходя из полномочий пользователя, делающего запрос, и полномочий, связанных с исходной информацией; наполнение базы поискового индекса; использование информации о структуре документа; агрегирование полученной информации из файлов в документы, при этом файлы рассматриваются как версии документа; объединение через корпоративную систему поиска информации версий файлов в документ для построения общего индекса;

нахождение через корпоративную систему поиска информации документов с похожими структурными данными из мета-информации для последующего вывода найденной информации пользователю при наличии у него необходимых полномочий; объединение по данным признакам

через корпоративную систему поиска информации этих документов в более верхнеуровневые структуры, такие как "проект", "направление"; отображение при проведении поиска результатов в виде "проекта", "направления". 2 н.п. ф-лы, 1 ил.



RU 2729224 C2

RU 2729224 C2



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC

G06F 16/338 (2019.08); *G06F 16/31* (2019.08); *G06F 16/93* (2019.08)(21)(22) Application: **2019100812, 10.01.2019**(24) Effective date for property rights:
10.01.2019Registration date:
05.08.2020

Priority:

(22) Date of filing: **10.01.2019**(43) Application published: **10.07.2020 Bull. № 19**(45) Date of publication: **05.08.2020 Bull. № 22**

Mail address:

**660077, g. Krasnoyarsk, ul. Baturina, 30/3-80, OOO
"Maksiol", direktoru**

(72) Inventor(s):

**Bagaev Maksim Sergeevich (RU),
Bagaeva Olga Lvovna (RU)**

(73) Proprietor(s):

**Obshchestvo s ogranichennoj otvetstvennostyu
"MAKSIOL" (RU)**(54) **INFORMATION RETRIEVAL METHOD AND CORPORATE INFORMATION RETRIEVAL SYSTEM**

(57) Abstract:

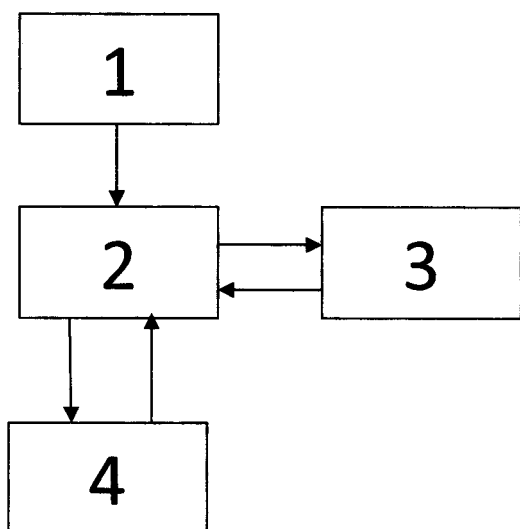
FIELD: computer equipment.

SUBSTANCE: invention relates to computer engineering. Disclosed is an information retrieval method implemented in a corporate information retrieval system configured to extract results from a plurality of different sources, comprising steps of receiving a request at an input; at the output, this request is issued in a set of different sources, wherein it is characterized by a hierarchical multi-tree list view of search results in form of one or more sets of related graphs, which nodes do not contain cycles, but can have duplicates and intersections with other sets or branches, wherein there can be a non-unique path, characterized by a sequence of nodes, from root to node, tree roots have a list representation and can be opened if necessary or interacting with the user interactively; includes execution of search requests; performing search query analysis by means of morphological analysis; accessing a search index database through a corporate information retrieval system and extracting search results therefrom; generating search results using a corporate information search system based on the right of the user making the

request, and rights associated with the source information; filling a search index base; using information on document structure; aggregation of obtained information from files into documents, wherein files are considered as version of document; merging, via a corporate information search system, versions of files into a document for constructing a common index; finding, through a corporate information retrieval system, documents with similar structural data from meta-information for subsequent output of found information to a user in the presence of necessary rights; combining these documents through the corporate information retrieval system into higher-level structures, such as "project", "direction"; display when searching results in the form of "project", "direction".

EFFECT: technical result is possibility of searching information with provision of information aggregation, information presentation form and compliance with information security requirements when providing access to documents.

2 cl, 1 dwg



Изобретение предназначено для крупных и средних предприятий, с большим объемом документов в информационной системе (ИС) предприятия и большим числом сотрудников.

Особую актуальность корпоративная система поиска информации (КСПИ) приобретает в компаниях имеющих разветвленную организационную структуру в разных регионах страны или земного шара.

Корпоративный поиск – это функция, которая должна быстро и просто приводить пользователя к небольшому количеству результатов из миллионов документов. Это задача сильно отличается от поиска в Интернет, осуществляемого по огромному массиву информации.

Известна Система управления информацией (патент GB2338324, МПК G06F17/30, оп. 15.12.1999), используемая при поиске данных из множества блоков данных, такая как сборник научных статей, обеспечивает отображение группы единиц данных в виртуальном пространстве и указание каждым блоком данных его корреляции с критерием поиска. Между различными блоками данных предоставляются ссылки, которые помогают перемещаться по данным во время поиска. Кроме того, может быть дано указание на предыдущий или текущий поиск другим поисковым устройством, на котором будет отображаться путь поиска пользователей через группу единиц данных.

Недостаток известного технического решения заключается в том, что построены только связи документов, не определяя их версию.

Известен Компьютеризированный метод (патент US6738678, МПК G06F17/30, оп. 18.05.2004), который определяет ранжирование документов, включая информационный контент. В настоящем методе используется анализ контента и связности. Входной набор документов представляется в виде графа окрестности в памяти. На графике каждый узел представляет один документ, и каждый направленный фронт, соединяющий пару узлов, представляет собой связь между двумя документами. Входной набор документов, представленных на графике, оценивается в соответствии с содержимым документов. Подмножество документов выбирается из набора входных данных, если ранжирование содержимого выбранных документов превышает первый заданный порог. Узлы, представляющие любые документы, отличные от выбранных документов, удаляются из графика. Выбранный поднабор документов оценивается в соответствии с привязкой документов, а выходной набор документов, превышающий второй заданный порог, выбирается для представления пользователям.

Наиболее близким техническим решением, принятым в качестве прототипа является СПОСОБ, СИСТЕМА И КОМПЬЮТЕРНЫЙ ПРОГРАММНЫЙ ПРОДУКТ ДЛЯ ПОИСКА, НАВИГАЦИИ И РАНЖИРОВАНИЯ ДОКУМЕНТОВ В ПЕРСОНАЛЬНОЙ СЕТИ (Патент РФ 2388050, МПК G06F17/30, оп. 27.04.2010). В способе определяют ненаправленную взвешенную связь между двумя из упомянутого множества документов на основе подобия; определяют направленную взвешенную связь между, по меньшей мере, двумя из упомянутого множества документов; добавляют упомянутые определенные ненаправленные взвешенные связи к упомянутым определенным направленным взвешенным связям для создания гибридной сети, имеющей связи, и выполняют алгоритм анализа связей, получающий связи упомянутой гибридной сети в качестве своих входных данных, причем упомянутый алгоритм включает в себя, по меньшей мере, анализ связей вперед и анализ связей назад. Выходными данными упомянутого алгоритма является набор оценок в анализе связей.

Недостаток заключается в том, известное техническое решение является универсальным и касается только построения связей между документами, но

отсутствуют связи между версиями документов.

Задача изобретения заключается в разработке способа агрегирования информации, формы представления и соответствия требованиям информационной безопасности (ИБ) при предоставлении доступа к документам.

5 Техническим результатом от применения изобретения является предоставление возможности поиска информации с обеспечением агрегирования информации, формы представления информации и соответствия требованиям информационной безопасности (ИБ) при предоставлении доступа к документам.

10 Это достигается за счет того, что способ извлечения информации, реализуемый в корпоративной системе поиска информации, выполненной с возможностью извлечения результатов из совокупности различных источников, содержащий этапы, на которых на входе принимают запрос; на выходе выдают этот запрос в совокупности различных источников, **согласно изобретению, характеризуется** иерархическим мульти-
15 - древовидном списочном **представлении** результатов поиска в виде одного или многих наборов связанных графов, узлы которых не содержат циклы, но могут иметь дубли и пересечения с другими наборами или ветвями, при этом могут иметь неunikальный путь, характеризующийся последовательностью узлов, от корня до узла, корни деревьев имеют списочное представление, и могут раскрываться при необходимости или при взаимодействии с пользователем интерактивно. Способ **включает выполнение** поисковых
20 запросов посредством веб-интерфейса, мобильного приложения, или в иной системе посредством программного интерфейса приложения API, предоставляющего услуги, как основной интерфейс пользователя, может быть использована поисковая строка; далее **проведение разбора** поискового запроса при помощи морфологического анализа, используя базу или иное хранилище похожих слов или их компонентов, для расширения
25 возможности поиска; применяя корпоративную систему поиска информации; **обращение** к базе поискового индекса через корпоративную систему поиска информации и извлечения из нее результатов поиска в виде ссылки, документа, проекта, файла, объекта, контактной информации для взаимодействия с владельцем информации; **формирование результатов поиска** используя корпоративную систему поиска информации, исходя из
30 полномочий пользователя, делающего запрос и полномочий связанных с исходной информацией; наполнение базы поискового индекса через взаимодействие корпоративной системы поиска информации с информационной системой компании и получение от нее информации для создания поискового индекса, на основании которого осуществляется поиск; **использование** информации о структуре документа для повышения
35 качества индексации, поиска и отображения документа; **агрегирование** полученной информации из файлов в документы, при этом файлы рассматриваются как версии документа; **объединение** через корпоративную систему поиска информации версий файлов в документ для построения общего индекса; **нахождение** через корпоративную систему поиска информации документов с похожими структурными данными из мета-
40 - информации для последующего вывода найденной информации пользователю при наличии у него необходимых полномочий; **объединение** по данным признакам через корпоративную систему поиска информации этих документов в более верхнеуровневые структуры, такие как "проект", "направление"; **отображение** при проведении поиска результатов в виде "проекта", "направления", в котором может быть указано
45 наименование "проекта", создатели, владельцы "проекта", "направления" и прочая информация доступная пользователю, согласно его полномочий.

Это достигается также за счет того, что корпоративная система поиска информации (КСПИ), содержит основные структуры, в виде агента источника данных (АИД),

системы управления Базой Данных (СУБД), системы агрегирования информации (САИ), системы интерфейса пользователя (СИП); агент источника данных АИД предназначен для поставки документов, файлов, извлечения ключевых слов, ранжирования ключевых слов, индексации, считывания свойств, полномочий или иной мета - информации, а также для использования технологии преобразования графического изображения текста в компьютерный текст с помощью алгоритма распознавания графических образов (OCR), формирования индексной информации, которая поступает в систему управления базой данных СУБД, агент источника данных АИД содержит набор фильтров, которые обеспечивают разбор различных форматов файлов, причем мета - информация поступает и от агента и из фильтра, из которого она извлекается, формирования поискового индекса для сохранения в системе управления базой данных СУБД, причем агент источника данных АИД может быть реализован на нескольких функциональных информационных системах (ИС); система управления базой данных СУБД предназначена для хранения: морфологических словоформ для расширения возможностей поиска, штатной или организационной структуры предприятия полученной из информационных систем ИС предприятия, полномочий, поискового индекса; система агрегирования информации САИ предназначена для проведения концентрирования отдельных потоков информации в единую структуру, с возможностью формирования индексной информации для объединения файлов в документы, документов в проекты; система интерфейса пользователя СИП предназначена для реализации взаимодействия с пользователем, посредством различных программных и аппаратных средств; при этом выход агента источника данных АИД соединен со входом системы управления базой данных СУБД, и система управления базой данных СУБД соединена с системой агрегирования информации САИ и с системой интерфейса пользователя СИП; агент источника данных АИД осуществляет взаимодействие с информационной системой, файловыми серверами, системой документооборота, интранет и интернет-сайтами, получает оттуда информацию и полномочия, агент источника данных АИД полностью считывает сам файл, выбирает из системы всю информацию о полномочиях, название файла, наименование файла, теги, флаги, то есть внешние атрибуты файла и отправляет вместе со всей мета - информацией, полномочия, дату создания, создателя, которую он взял из информационной системы ИС, в систему управления базой данных СУБД; агент источника данных АИД предназначен для обработки информации, содержит модуль распознавания (OCR), формирует индексную информацию, которую можно поместить в систему управления базой данных СУБД, содержит набор фильтров, которые обеспечивают разбор различных форматов файлов, причем мета - информация поступает и от агента и из фильтра, который ее извлекает; в систему управления базой данных СУБД поступает мета - информация, а также информация для создания поискового индекса; система управления базой данных СУБД является компонентом корпоративной системы поиска информации, которая хранит индексы, при поступлении запроса через систему интерфейса пользователя СИП, происходит обращение к системе управления базой данных СУБД и получение необходимой информации из системы управления базой данных СУБД, при этом выборка осуществляется на основании информации, которая храниться в системе управления базой данных СУБД с использованием морфологии, разбора запроса; агент источника данных АИД выполнен с возможностью сжимать и отправлять полученную информацию на сервер, где находится система управления базой данных СУБД, там информация разворачивается и вкладывается в систему управления базой данных СУБД, так как каналы связи могут быть слабыми; система интерфейса пользователя

СИП формирует запросы к системе управления базой данных СУБД для извлечения информации, с учетом морфологии, разбора слов, осуществляет сами запросы и уточнения, система интерфейса пользователя СИП осуществляет двухсторонний обмен информацией с системой управления базой данных СУБД.

5 **Новые существенные признаки данного изобретения:**

1. Иерархическое мульти-древовидное списочное представление результатов поиска, а именно в виде одного или многих наборов связанных графов, где узлы графа не содержат циклы, но могут иметь дубли и пересечения с другими наборами или ветвями, а также иметь не уникальный путь (последовательностью узлов) от корня до узла, при этом корни деревьев могут иметь списочное представление и раскрываться при необходимости или при взаимодействии с пользователем как результат поиска, использование дерева как способа интерактивно уточнять запрос поиска.

2. Агрегирование информации – то есть объединение версий документа/файлов в документ, объединение документов в проект, и так далее по доступной иерархии; возможно использование агрегированного индекса для поиска.

3. Последовательное агрегирование версий документа, в связи с накоплением отличий первоначальной версии к конечной.

4. Соответствие требованиям современной Информационной Безопасности – а именно управление полномочиями при поиске.

20 Существенные отличия заявляемого изобретения от известных технических решений.

1. Корпоративный поиск - поиск может работать с изначально закрытой информацией. В известных решениях интернет поиск - поиск для всех изначально открытой информации.

2. Работа с дублями - возможность консолидации всех версий, оперирование сущностью "Документ", который может в себя включать множество версий и множество файлов. В известных решениях работа с дублями - интернет поиск дубли исключает, они ищут самую первую версию или наиболее востребованную пользователями и считают ее наиболее релевантной.

3. Может проходить агрегирование информации (агрегирование документов в проекты, и далее по иерархии). Возможность использования иерархии, в пределах которой может рассматриваться поиск документов. В иерархии возможно использовать дополнительные свойства документа (дата создания, автор создания) и полномочия документа, которые могут предоставлять системы хранения документа, а также в КСПИ возможно выделение у документа - тегов (слово, словосочетание, которое может использоваться в поиске). В известных решениях выдаются различные документы, нет агрегации в проекты, может быть огромное количество документов, в которых пользователь может запутаться.

4. В системе не используется анализ релевантности на основании информации от пользователей. В известных решениях при поиске учитывается релевантность анализа документа на основании информации от пользователей. Накапливается статистика, сколько пользователей посмотрело и сколько времени на просмотр затрачено.

5. В поисковой системе может учитываться вся информация, все версии. В известных решениях на основании графов строят связи документов, чтобы выбрать необходимый документ, отбрасывая ненужные версии.

45 6. В КСПИ могут отсутствовать ссылки с документа на документ. Ранжирование по этим признакам также может отсутствовать. (Проект не может быть более популярным, если на него кто-то ссылается). КСПИ может работать с закрытой информацией. В известных решениях ранжируют при поиске документов на основании

ссылок документа на документ. Больше ссылок (больше пользователей посмотрело) - считается более релевантным.

7. В процессе поиска - могут вырисовываться две самостоятельные опции:

а. иерархическое мульти-древовидное списочное представление результатов поиска;

б. возможность управлять полномочиями, в соответствии с полномочиями возможно отображение или документа, или указывается средство, способ для связи с владельцем документа, чтобы была возможность получить сам документ (в результате поиска - документ или отдел разработавший документ - ФИО начальника отдела, телефон, e-mail). А также существует возможность запросить полномочия на разыскиваемый документ через корпоративные системы запроса полномочий (систему IDM и т.д.). В известных решениях найденная информация представлена в виде ссылок на документы.

8. КСПИ может оперировать с сущностью документа, то есть информация из всех версий может консолидироваться и использоваться для поиска этого документа.

Документы могут также не относиться к одному проекту, а быть разбросанными по разным системам хранения данных. В известных решениях информация не консолидируется, предоставляются пользователю различные варианты документов.

9. В КСПИ поиск может вестись не отдельного документа, а всего проекта целиком. В известных решениях ведется поиск отдельного документа.

10. Поисковый индекс может строиться по объекту - "документ" состоящего из множества файлов-версий документа. Индекс, возможно используемый в предлагаемой системе - агрегированный. Он может содержать в себе информацию из всех версий документа. (информация в разных версиях документа может принципиально отличаться). В известных решениях поисковый индекс - строится по одному файлу (он же документ).

11. При наличии объекта - "проект" поиск может проходить по множеству файлов и множеству версий. Есть возможность провести дополнительную агрегацию на верхний уровень - выше уровня документа. (например проект, проекты, каталог отдела). В известных решениях ведется поиск только отдельного документа, нет агрегации.

12. Удовлетворительный результат поиска - это предоставление всех версии документа, или проекта, или проекты, или наименование отдела, который этим занимается (является владельцем данной информации), в зависимости от полномочий запрашивающего лица. В известных решениях результат поиска - открытый документ. Данный поиск и его результаты не зависят от полномочий лица, осуществляющего поиск.

Сущность технического решения.

КСПИ предоставляет для пользователя возможность делать поисковые запросы посредством веб - интерфейса, мобильного приложения, или в иной системе посредством программного интерфейса приложения (API), предоставляющей услуги. Как основной интерфейс пользователя может быть использована поисковая строка.

Поисковая система предназначена для крупных и средних предприятий, с большим объемом документов в ИС предприятия и большим числом сотрудников. Особую актуальность КСПИ приобретает в Компаниях имеющих разветвленную организационную структуру в разных регионах страны или земного шара.

Новизна КСПИ заключается в агрегировании информации, форме представления и соответствии требованиям ИБ при предоставлении доступа к документам.

Результатом поиска может служить документ, проект, объект, верхнеуровневая структура ИС и/или контактная информация для взаимодействия с владельцем информации. КСПИ формирует результат поиска исходя из полномочий пользователя делающего запрос и полномочий связанных с исходной информацией.

Поисковый запрос может быть разобран при помощи морфологического анализа, может быть использована база или иное хранилище похожих слов или их компонентов, которые расширяют возможности поиска.

КСПИ взаимодействует с ИС Компании и получает от нее информацию для создания
5 поискового индекса на основании, которого осуществляется поиск.

С ИС Компании КСПИ может получать информацию о составе групп безопасности, считывать информацию о полномочиях с самих вложенных файлов, директорий или иных структур хранения данных.

КСПИ также может получать информацию о штатной структуре предприятия,
10 контактную информацию, и другую дополнительную информацию в виде тегов, флагов самого документа, свойств документа и прочей мета - информации для последующего вывода найденной информации пользователю при наличии у него необходимых полномочий.

Использование информации о структуре документа, может использоваться для
15 повышения качества индексации, поиска и отображения документа.

Полученная информация о файлах может агрегироваться в документы. КСПИ может объединять версии и файлы в документ, чтобы построить общий индекс. Однако возможен вывод и отдельных документов.

Так же КСПИ может находить документы с похожими структурными данными
20 (заголовки, теги, дата создания, прочая мета - информация), по данным признакам КСПИ может объединять эти документы в "проект", при проведении поиска результат может отображаться в виде "проекта". В нем может быть указано наименование проекта, создатели, владельцы "проекта" и прочая информация доступная пользователю, согласно его полномочий. Аналогичными методами проекты могут объединяться в более
25 высокоуровневые структуры.

Для ясного понимания, данное описание касается упрощенных вариантов осуществления настоящего изобретения. Многие варианты осуществления настоящего технического решения будут обладать гораздо большей сложностью.

Предоставление информации Пользователю.

30 После отправки поискового запроса (в запросе может быть одно или несколько слов), пользователь может получить ответ в агрегированном виде, в виде свернутых иерархических структур (деревьев). Корнями являются структуры верхнего уровня, например Государства или Регионы, в которых представлена компания. По мере развертывания дерева могут открыться такие группы как:

- 35 а) регион;
- б) организации;
- с) подразделения;
- д) проекты, структуры и пр.
- е) документ/ы;
- 40 ф) версии документа.

Данный список не является полным или исключительным, возможно представление в дереве других групп, специалисты в данной области могут создавать другие группы, остающиеся в границах объема настоящей технологии. Список может формироваться на основании статистической информации в автоматическом режиме. Данный список
45 групп не должен интерпретироваться как единственный вариант осуществления этого элемента настоящей технологии и/или то, что было описано выше, является единственным вариантом осуществления этого элемента настоящей технологии.

Пользователь при развертывании дерева может дойти:

1) до проекта, документа и версий документа при наличии у него полномочий;
 2) до каталога отдела (подразделения), при отсутствии у него полномочий, с предоставлением пользователю информации о:

- a) подразделении владельце;
- b) руководителе подразделения;
- c) владельце ресурса;
- d) контактной информации;

Данный список не является полным или исключительным, возможно предоставление другой информации.

Данный список информации не должен интерпретироваться как единственный вариант осуществления этого элемента настоящей технологии и/или то, что было описано выше, является единственным вариантом осуществления этого элемента настоящей технологии.

На основании полученной информации пользователь может запросить необходимые ему полномочия, к примеру:

- ФИО и должность могут подсказать, к кому необходимо обратиться за необходимыми полномочиями. Телефон или e-мейл может быть предоставлен для запросов полномочий.

При наличии интегрированной системы управления учетными данными (IDM), пользователь может запросить полномочия на просмотр документов в автоматическом режиме.

Возможно, что с запрашиваемыми документами (версиями, файлами, документами, проектами, объектами) ведется работа в других подразделениях, данная информация также может отобразиться в дереве при наличии такой информации и полномочий пользователя для ее получения. В иерархии версий может отобразиться, что с данным проектом (версией, файлом, документом, объектом) работал сторонний отдел (или сторонние отделы) и результаты работы другого отдела, возможно, будут храниться в другом каталоге, ресурсе или ИС. В зависимости от наличия у пользователя необходимых полномочий, он сможет увидеть полную развернутую информацию. Если же нет полномочий, он сможет запросить необходимую информацию и у стороннего отдела, используя контактную информацию или запросить полномочия через систему IDM, при ее наличии в Компании. Реализация управления безопасностью является опциональной и основывается на специфике заказчика.

На рисунке показана схема работы КСПИ.

КСПИ состоит из следующих основных структур: АИД 1, СУБД 2, САИ 3, СИП 4. Все примеры и используемые в настоящем описании условные конструкции предназначены для того, чтобы можно было понять принцип настоящего технического решения, а не для установления границ ее объема. Специалисты в данной области могут разработать различные схемы, отдельно не описанные и не показанные в данном описании, но которые воплощают собой принципы настоящей технологии и находятся в границах ее объема.

Агент источника данных - АИД (1) - может осуществлять:

- a) поставку документов (файлов);
- b) парсинг (разбор) текстовой информации в том числе и OCR преобразование документов;
- c) экстракцию (извлечение) ключевых слов;
- d) ранжирование ключевых слов;
- f) считывание свойств, полномочий или иной мета - информации;

е) формирование поискового индекса для сохранения в СУБД

Данный список функционала АИД не должен интерпретироваться как единственный вариант осуществления этого элемента настоящей технологии и/или то, что было описано выше, является единственным вариантом осуществления этого элемента

5 настоящей технологии. АИД может быть реализован на нескольких функциональных ИС с различным функциональным разделением.

Система Управления Базой Данных СУБД (2) это структура, в которой осуществляется хранение:

а) морфологических словоформ для расширения возможностей поиска;

10 б) штатной или организационной структуры предприятия полученной из ИС предприятия;

в) полномочий;

г) поискового индекса.

Данный список хранимой в СУБД информации не должен интерпретироваться как

15 единственный вариант осуществления этого элемента настоящей технологии и/или то, что было описано выше, является единственным вариантом осуществления этого элемента настоящей технологии.

Система Агрегирования Информации САИ (3) это структура, в которой осуществляется анализ отдельных блоков информации с целью выявления общностей

20 и формирования поисковой информации для объединения файлов в документы, документов в проекты.

Система Интерфейса Пользователя СИП (4) может реализовать взаимодействие с пользователем, посредством различных программных и аппаратных средств.

Все заявленные здесь принципы, аспекты и варианты осуществления настоящего

25 изобретения, равно как и конкретные примеры, предназначены для обозначения их структурных и функциональных основ, вне зависимости от того, известны ли они на данный момент или будут разработаны в будущем.

Взаимодействие компонентов системы:

1.Агент источника данных (АИД), который осуществляет взаимодействие с

30 информационной системой (файловый сервер, система документооборота, интранет и интернет-сайты)(все что содержит в себе файлы), получает оттуда информацию, полномочия. АИД полностью считывает сам файл, берет из информационной системы всю информацию о полномочиях, название документа/файла, наименование документа/файла, теги, флаги, то есть внешние атрибуты файла. АИД содержит набор фильтров,

35 которые обеспечивают разбор различных форматов файлов/документов. Фильтры извлекают как и текстовую информацию, так и мета-информацию. Например - документ Microsoft Word содержит в себе информацию: сам текст, разметку, заголовки, и другую мета - информацию (кто создатель, дата создания и т.д.). Данный документ/файл является контейнером, содержащим различную информацию. Фильтр является средством

40 унификации на уровне парсинга различных форматов данных. Список фильтров может расширяться по мере необходимости осуществлять обработку новых форматов данных. Вся полученная информация из АИД поступает в СУБД для формирования поискового индекса. АИД может содержать в себе модуль распознавания текста в графической информации (OCR).

45 2. Система управления Базой Данных (СУБД) - это компонент системы, который хранит информацию необходимую для осуществления поиска. При поступлении запроса через СИП, происходит обращение к системе управления базой данных СУБД и получение необходимой информации из системы управления базой данных СУБД.

Выборка осуществляется на основании информации, которая храниться в системе управления базой данных СУБД с использованием морфологии, разбора запроса. Возможна ситуация когда полученную информацию АИД сжимает и отправляет на сервер, где находится СУБД, там информация разворачивается и вкладывается в систему управления базой данных СУБД, так как каналы связи могут быть слабыми. Может быть и другая архитектура. Сжатие - опционально, может быть, может не быть.

3. САИ осуществляет взаимодействие с системой управления базой данных СУБД и занимается анализом поисковой информации с целью выявления общностей у хранимой информации, выявление общностей позволяет объединять версии документов в документы, документы в проекты двигаясь вверх. Объединение в более верхнеуровневые структуры не является обязательным.

4. СИП (веб интерфейс, API и т.д.) - с помощью которого может взаимодействовать пользователь и/или иная система. Пользовательский интерфейс формирует запросы к системе управления базой данных СУБД для извлечения информации. С учетом морфологии. разбора слов осуществляет сами запросы и уточнения. Интерфейс осуществляет двухсторонний обмен информацией с системой управления базой данных СУБД.

Примеры работы поиска

Пример 1: поиск в корпоративной сети

1. Пользователь вводит в поисковую строку любое слово или словосочетание относящееся к интересующей его теме. Например - он может ввести слово - план 2019 года. Далее пользователь запускает в работу поисковую систему.

2. Интерфейсная часть посылает запрос на поисковый сервер или группу серверов, на которых хранится информация.

3. Начинается морфологический разбор запроса поиска, то есть используется база или иное хранилище подобных слов, которые расширяют возможности поиска.

4. Расширенный запрос поступает в базу, и оттуда достаются нужные документы или разные версии разыскиваемого документа в виде файлов.

5. Пользователю предоставляется информация в агрегированном виде, в виде свернутых иерархических структур (деревьев). Каждая структура начинается с глобальной группы, объединяющую верхнеуровневую группу документов, к примеру по региональному, организационному или иному общему признаку (Регион или Головная Организация), который является корневым.

6. Глобальный признак общий для большинства документов в этой структуре, затем дерево открывается дальше - ветви дерева - ветви иерархии образованные общими признаками с меньшим весом объектов под ними, конечными листьями дерева являются документы или версии документов. Как пример - первым уровнем от корня является - каталог - Регион или Головная Организация. В дальнейшем дерево открывается до следующих уровней - это будет - Организация, затем - отделы, потом - проекты, структуры и т.п., в конце документы или версии документов.

7. Отличительной особенностью системы поиска может быть представление результатов поиска в соответствии с полномочиями лица, которое запрашивает информацию. То есть у пользователя может быть, а может и не быть полномочий для просмотра запрашиваемых им документов или организационных групп.

8. Если у пользователя есть необходимые полномочия (программа поиска получает информацию об полномочиях каждого конкретного пользователя из информационной системы предприятия, в которой работает данный пользователь) поисковая система представит ему информацию в агрегированном виде, а пользователь, получив

информацию в свернутом виде, сможет ее развернуть (получая дополнительную информацию на каждом уровне) до проекта, до документа и до конкретной версии разыскиваемого документа.

5 9. Если же у пользователя нет достаточных полномочий для просмотра данного документа, программа поиска предоставляет ему информацию также в агрегированном виде, но пользователь при разворачивании уровней сможет получить только общую информацию о разыскиваемом им документе. Система может предоставить контактные данные о владельце (создателе, руководителе - занимающего максимальную должность) этого подразделения или отдела, в котором создавался или находится интересующий

10 пользователя документ.

10. Пользователь может использовать полученную контактную информацию (к примеру ФИО владельца, должность, телефон, e-мэйл), предоставленную системой поиска для получения доступа к запрашиваемому документу. Например, позвонить по контактному телефону, написать на e-мэйл владельца документа. При применении на

15 предприятии системы автоматизации управления полномочиями - пользователь сможет запросить полномочия в автоматическом режиме. Пользователь также может запросить полномочия на поиск – т.е. возможность искать в данном каталоге/ресурсе согласовав это с владельцем ресурса. Полномочия на поиск не означают возможности читать документы.

20 11. Возможна и такая ситуация, что с разыскиваемым документом проводится работа в других отделах или предприятиях (в крупном холдинге, объединенном в единую информационную сеть). Данная информация также отобразится в результате поиска документа. В иерархии версий может отобразиться какой отдел работал или работает с какой версией документа и где и у кого храниться на данный момент версия документа.

25 Если есть полномочия на просмотр - то покажется вся информация, если нет - то только информация у кого можно запросить необходимые полномочия.

12. Опция проверка полномочий может быть в системе поиска информации, а может и не быть. В зависимости от пожеланий заказчика. В случае если данной опции не будет - результат поиска можно будет открыть до конечной информации - до версий

30 документов.

Пример 2 (Интернет поиск).

Пользователь запрашивает информацию о программном обеспечении – к примеру об игре. Так как пользователь ленив, то он пишет в запросе просто название игры и только (далее “ИГРА”). Система возвращает ему результат в виде нескольких деревьев,

35 корнями являются типовые уточнения запроса с этой игрой:

- обзоры “ИГРА”
- прохождение “ИГРА”
- приобретение “ИГРА”
- чит-коды “ИГРА”
- 40 - скачать “ИГРА”

и т.д. ранжируя по популярности запросов.

Пользователь сразу видит, что ему необходимо уточнить, что именно он хочет получить. Он может в начале раскрыть дерево “обзоры” и увидеть там:

- обзоры на русском;
- 45 - обзоры на английском;
- обзоры на других языках;

Т.е. в принципе очевидные варианты уточнения.

К примеру, он выбирает – обзоры на русском и он получает типовой ответ поисковика

с таким уточнением.

Затем пользователь выбирает “скачать” и получает различные источники, где эту игру предлагают скачать сгруппированные по типам:

- торрент - трекеры;
- сайты официальной продажи;
- FTP сервера;
- и т.д. и т.п.

Пример 3 (интернет поиск).

Пользователь интересуется здоровьем и делает запрос по названию болезни (далее “Болезнь”).

В результате запроса он получает типовой набор уточнений:

- диагностика “Болезнь”;
- лечение “Болезнь”;
- симптомы “Болезнь”;
- фото “Болезнь”;

и т.д.

Предположим пользователь уже знает какая именно “Болезнь” его интересует и хочет получить информацию о современных методах лечения – для этого он начинает раскрывать дерево “лечение”.

Далее он получает:

- традиционные методики лечения “Болезнь”;
- препараты для лечения “Болезнь”;
- медицинские центры лечения “Болезнь”;
- зарубежные клиники лечения “Болезнь”;
- народные методики лечения “Болезнь”;

Соответственно он сразу видит картину и может ориентироваться и принимать решение куда ему идти, а не открывать каждый сайт на первой странице в надежде найти нужную ему информацию.

К примеру он выбрал препараты – в ответ он получает весь список препаратов.

А уже открывая нужный, он может получить:

- описание;
- применение;
- приобретение;

Т.е. все что нужно – прочитал про применение – раскрыл ветку приобретение и заказал препарат в магазине, предварительно выбрав магазин в своем городе (а не перебирая все магазины, или не делая отдельный поиск на каждую ветку).

Пример 4 (интернет поиск).

Пользователь не знает, что он ищет и хочет узнать, что же это – для показа сложности задачи представим, что иностранец хочет понять значение слова “Коса” в Русском языке. Он делает запрос “Коса” и соответственно система поиска должна вернуть ему все варианты включая исправления:

- Коса (причёска) — волосы, сплетённые между собой в длину;
- Коса́ — сельскохозяйственный ручной носимый инструмент для скашивания травы;
- Коса – отмель на реке;
- Коса (провода) — пучок проводов в автомобиле, связывающий двигатель, коробку передач и компьютер;
- Коза - животное;

и т.д.

Т.е. пользователь сразу поймет, с чем он имеет дело и куда можно уточнять запрос
Из примеров видно, что интерактивное уточнение вопросов в пределах
иерархического мульти-древовидного списочного представления результатов поиска
очень удобен пользователю и будет актуален особенно для мобильных пользователей
5 где ввод дополнительной текстовой информации затруднен.

В интернет поиске нет исходной навязанной иерархии, но она может быть
сформирована на основе уже имеющейся типовой информации исходя из статистической
информации о поиске.

Данный способ может позволить уточнять запрос поиска пользователем в
10 интерактивном режиме (в реальном времени) или являться статической структурой,
раскрываемой в реальном времени.

Пользователь может дойти до конечного документа на сайте - то есть в конце
интерактивного поиска выводится ссылка на страница сайта с разыскиваемой
информацией.

15

(57) Формула изобретения

1. Способ извлечения информации, реализуемый в корпоративной системе поиска
информации, выполненной с возможностью извлечения результатов из совокупности
различных источников, содержащий этапы, на которых на входе принимают запрос;
20 на выходе выдают этот запрос в совокупности различных источников, отличающийся
тем, что он характеризуется иерархическим мульти-древовидным списочным
представлением результатов поиска в виде одного или многих наборов связанных
графов, узлы которых не содержат циклы, но могут иметь дубли и пересечения с другими
наборами или ветвями, при этом могут иметь неуникальный путь, характеризующийся
25 последовательностью узлов, от корня до узла, корни деревьев имеют списочное
представление и могут раскрываться при необходимости или при взаимодействии с
пользователем интерактивно; включает выполнение поисковых запросов посредством
веб-интерфейса, мобильного приложения, или в иной системе посредством программного
интерфейса приложения API, предоставляющего услуги, как основной интерфейс
30 пользователя, может быть использована поисковая строка; далее проведение разбора
поискового запроса при помощи морфологического анализа, используя базу или иное
хранилище похожих слов или их компонентов, для расширения возможности поиска,
применяя корпоративную систему поиска информации; обращение к базе поискового
индекса через корпоративную систему поиска информации и извлечение из нее
35 результатов поиска в виде ссылки, документа, проекта, файла, объекта, контактной
информации для взаимодействия с владельцем информации; формирование результатов
поиска, используя корпоративную систему поиска информации, исходя из полномочий
пользователя, делающего запрос, и полномочий, связанных с исходной информацией;
наполнение базы поискового индекса через взаимодействие корпоративной системы
40 поиска информации с информационной системой компании и получение от нее
информации для создания поискового индекса, на основании которого осуществляется
поиск; использование информации о структуре документа для повышения качества
индексации, поиска и отображения документа; агрегирование полученной информации
из файлов в документы, при этом файлы рассматриваются как версии документа;
45 объединение через корпоративную систему поиска информации версий файлов в
документ для построения общего индекса; нахождение через корпоративную систему
поиска информации документов с похожими структурными данными из мета-
информации для последующего вывода найденной информации пользователю при

наличии у него необходимых полномочий; объединение по данным признакам через корпоративную систему поиска информации этих документов в более верхнеуровневые структуры, такие как "проект", "направление"; отображение при проведении поиска результатов в виде "проекта", "направления", в котором может быть указано наименование "проекта", создатели, владельцы "проекта", "направления" и прочая информация, доступная пользователю согласно его полномочиям.

2. Корпоративная система поиска информации (КСПИ) для осуществления способа по п.1, содержащая основные структуры в виде агента источника данных, системы управления базой данных, системы агрегирования информации, системы интерфейса пользователя; агент источника данных предназначен для поставки документов, файлов, извлечения ключевых слов, ранжирования ключевых слов, индексации, считывания свойств, полномочий или иной мета-информации, а также для использования технологии преобразования графического изображения текста в компьютерный текст с помощью алгоритма распознавания графических образов, формирования индексной информации, которая поступает в систему управления базой данных; агент источника данных содержит набор фильтров, которые обеспечивают разбор различных форматов файлов, причем мета-информация поступает и от агента и из фильтра, из которого она извлекается, формирования поискового индекса для сохранения в системе управления базой данных, причем агент источника данных может быть реализован на нескольких функциональных информационных системах; система управления базой данных предназначена для хранения: морфологических словоформ для расширения возможностей поиска, штатной или организационной структуры предприятия полученной из информационных систем предприятия, полномочий, поискового индекса; система агрегирования информации предназначена для проведения концентрирования отдельных потоков информации в единую структуру; с возможностью формирования индексной информации для объединения файлов в документы, документов в проекты; система интерфейса пользователя предназначена для реализации взаимодействия с пользователем посредством различных программных и аппаратных средств; при этом выход агента источника данных соединен со входом системы управления базой данных, и система управления базой данных соединена с системой агрегирования информации и с системой интерфейса пользователя; агент источника данных осуществляет взаимодействие с информационной системой, файловыми серверами, системой документооборота, интранет и интернет-сайтами, получает оттуда информацию и полномочия, агент источника данных полностью считывает сам файл, выбирает из системы всю информацию о полномочиях, название файла, наименование файла, теги, флаги, то есть внешние атрибуты файла и отправляет вместе со всей мета-информацией, полномочия, дату создания, создателя, которую он взял из информационной системы, в систему управления базой данных; агент источника данных предназначен для обработки информации, содержит модуль распознавания, формирует индексную информацию, которую можно поместить в систему управления базой данных, содержит набор фильтров, которые обеспечивают разбор различных форматов файлов, причем мета-информация поступает и от агента и из фильтра, который ее извлекает; в систему управления базой данных поступает мета-информация, а также информация для создания поискового индекса; система управления базой данных является компонентом корпоративной системы поиска информации, которая хранит индексы, при поступлении запроса через систему интерфейса пользователя, происходит обращение к системе управления базой данных и получение необходимой информации из системы управления базой данных; при этом выборка осуществляется на основании информации, которая

храниться в системе управления базой данных с использованием морфологии, разбора запроса; агент источника данных выполнен с возможностью сжимать и отправлять полученную информацию на сервер, где находится система управления базой данных, там информация разворачивается и вкладывается в систему управления базой данных, так как каналы связи могут быть слабыми; система интерфейса пользователя формирует запросы к системе управления базой данных для извлечения информации, с учетом морфологии, разбора слов, осуществляет сами запросы и уточнения, система интерфейса пользователя осуществляет двухсторонний обмен информацией с системой управления базой данных.

10

15

20

25

30

35

40

45

