



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ПОЛЕЗНОЙ МОДЕЛИ К ПАТЕНТУ

(21)(22) Заявка: 2011152639/08, 23.12.2011

(24) Дата начала отсчета срока действия патента:
23.12.2011

Приоритет(ы):

(22) Дата подачи заявки: 23.12.2011

(45) Опубликовано: 27.06.2012 Бюл. № 18

Адрес для переписки:

119415, Москва, Вернадского пр-кт, 41, стр.1,
генеральному директору ООО "Газпром
газнадзор" В.Н. Медведеву

(72) Автор(ы):

Медведев Вячеслав Николаевич (RU),
Тухбатуллин Фарит Гарифович (RU),
Докутович Алексей Борисович (RU),
Шапиро Владимир Дмитриевич (RU),
Андреев Алексей Евгеньевич (RU),
Пономарев Владимир Николаевич (RU),
Лим Владимир Григорьевич (RU),
Кац Илья Дмитриевич (RU),
Коваленко Сергей Владимирович (RU),
Винокурцев Александр Георгиевич (RU),
Воеводин Илья Геннадьевич (RU)

(73) Патентообладатель(и):

Общество с ограниченной ответственностью
"Газпром газнадзор" (ООО "Газпром
газнадзор") (RU)

(54) СИСТЕМА СТРУКТУРИРОВАННОГО ХРАНЕНИЯ НОРМАТИВНО-ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ В БАЗЕ ДАННЫХ

Формула полезной модели

1. Система структурированного хранения нормативно-технической документации в базе данных, содержащая блок интерфейса администратора, блок ввода документов в базу данных, блок хранения полных текстов документов, блок хранения фактографической информации о документах, блок первичной обработки полных текстов документов, блок хранения кодов и представлений терминов (понятий, концептов), блок хранения позиций терминов, блок хранения частот терминов, блок интерфейса информационно-поисковой системы, отличающаяся тем, что система дополнительно содержит блок интерфейса системы поддержки принятия решений, блок многомерной классификации документов с произвольным числом классификационных измерений n больше 3, блок классификаторов с количеством классификаторов многомерного классификационного пространства документов n больше 3, при этом первый вход блока многомерной классификации документов связан с выходом блока хранения фактографической информации о документах, остальные входы блока многомерной классификации документов со второго по $n+1$ вход связаны с выходами n классификаторов по соответствующим измерениям многомерного классификационного пространства документов, образующих блок классификаторов, выход блока многомерной классификации документов соединен с первым входом блока интерфейса системы поддержки принятия решений, выход блока интерфейса информационно-поисковой системы связан со вторым входом блока интерфейса системы поддержки

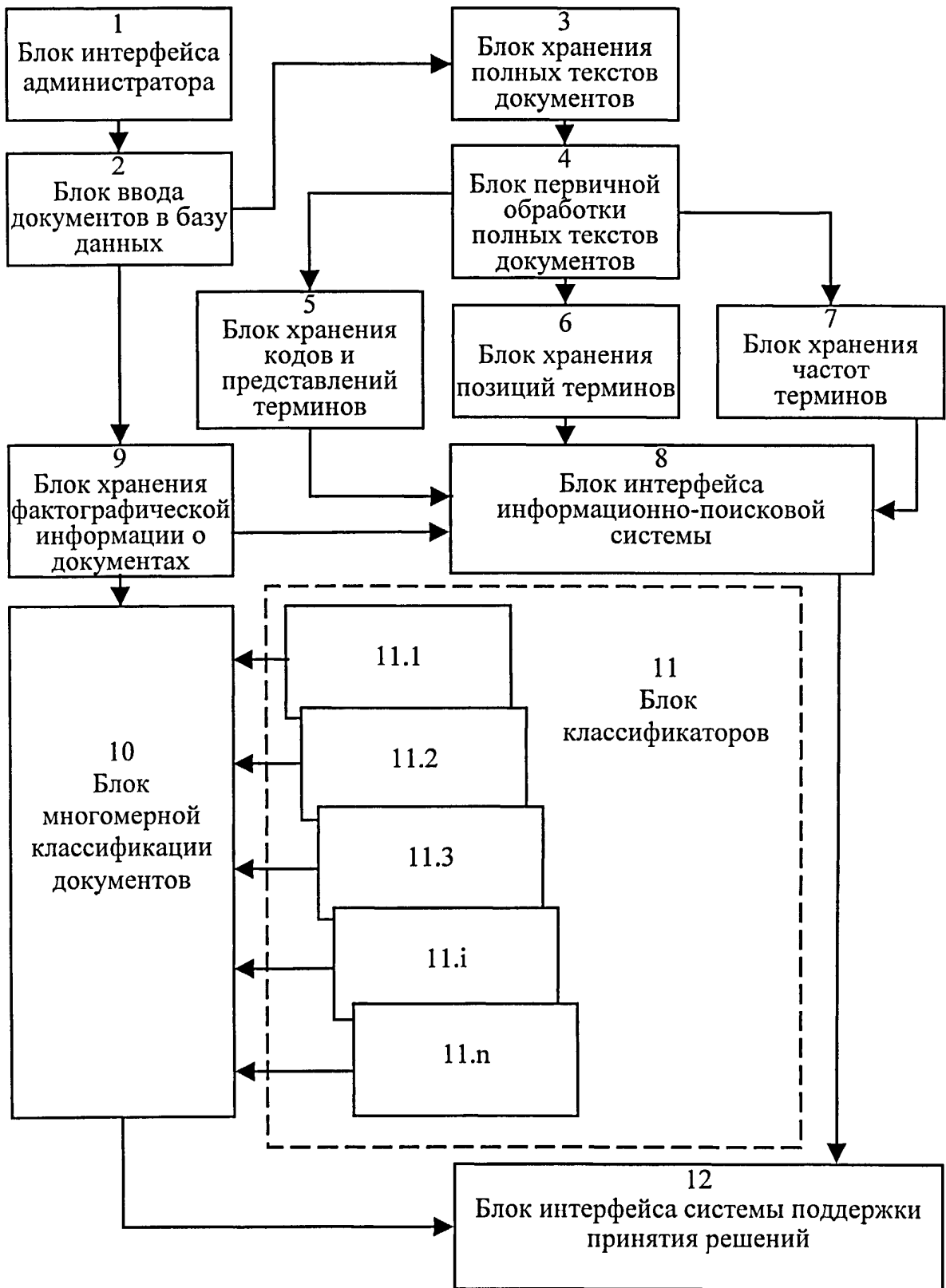
принятия решений.

2. Система по п.1, отличающаяся тем, что часть классификаторов сформирована автоматически путем анализа содержания обучающей выборки, формирования профилей классов обучающей выборки, анализа содержания полных текстов хранимых документов, автоматического сопоставления документов профилю класса на основе меры близости для их тематической классификации.

3. Система по п.1, отличающаяся тем, что блок интерфейса системы поддержки принятия решений состоит из модуля сопоставления объектов контроля и области контроля, модуля сопоставления контрольных вопросов и объекта контроля, модуля сопоставления фрагментов нормативно-технических документов и контрольного вопроса.

R U 1 1 7 6 6 6 U 1

R U 1 1 7 6 6 6 U 1



Полезная модель относится к вычислительной технике и может быть использована в промышленности, на транспорте, на предприятиях связи и в других отраслях экономики, охватываемых техническим регулированием, для повышения упорядоченности хранения нормативно-технических документов, представленных в электронно-цифровой форме, увеличения быстродействия и точности поиска документов; обеспечения возможностей извлечения фактов, описывающих техногенные объекты, формирования новых знаний о техногенных объектах на основе имеющихся, принятия решений на основе анализа нормативной информации; упрощения и оптимизации доступа потребителей к нормативно-технической информации.

Известна справочная правовая система хранения и поиска данных (патент RU 2223537 от 22.11.01 г., МПК⁷ G06F 17/30, 17/40), содержащая блок выбора вида поиска, формирователь запроса, один выход которого связан с блоком проведения поиска, вход-выход которого соединен через соответствующие шины с базами данных системы, блоком отображения и контроллером, предназначенным для управления поиском данных, в систему введены блок выбора условий поиска, вход которого связан с выходом блока выбора вида поиска, первый выход которого связан со входом формирователя запроса, второй выход соединен с входом формирователя атрибутов фильтра, первый вход-выход которого связан со вторым входом-выходом формирователя запроса, второй вход-выход - с блоком памяти атрибутов фильтра, первый вход-выход формирователя запроса соединен с блоком памяти запросов.

Базами данных системы являются база данных правовой информации и/или база данных экономической информации, база данных энциклопедии ситуаций и/или база данных толкового словаря, и/или база данных последних изменений в законодательстве.

При этом блок выбора вида поиска выполнен с возможностью навигации по базам данных с поэтапным представлением отклассифицированных списков запрашиваемых документов и/или фрагментов документов. Система также может включать средство сортировки документов, связанное с памятью, указанное средство и память подключены через соответствующие шины системы к соответствующим блокам системы.

Средство сортировки документов предназначено для сортировки документов по дате издания документов или по юридической силе.

Система включает также средство поиска документов по заданному числу сходных признаков, связанное по соответствующим шинам с соответствующими блоками системы.

Недостатком этой системы является то, что она не обеспечивает построение индекса массива документов, что снижает быстродействие системы, а также не гарантирует полноту и точность поисковой выдачи.

Известна также вычислительная система для интеллектуального анализа данных - хранилище данных для основанной на знаниях системы извлечения информации из данных (патент RU 2297665 от 28.04.03 г., МПК⁸ G06F 17/30), включающая в себя, по меньшей мере, одно хранилище данных, механизмы анализа более низкого уровня, механизмы анализа более высокого уровня, индексатор.

Система извлечения информации из данных, включающая в себя: по меньшей мере, одно хранилище, или склад данных, в котором содержатся объекты; по меньшей мере, один механизм анализа более низкого уровня, связанный с хранилищем данных и генерирующий выходные данные на основе первого набора правил, реализованных в указанном механизме анализа более низкого уровня; и, по меньшей мере, один механизм анализа более высокого уровня, принимающий выходные данные механизма анализа более низкого уровня и генерирующий свои выходные данные на основе второго набора

правил, реализованных в указанном механизме анализа более высокого уровня, причем выходные данные механизмов анализа более низкого и более высокого уровней присоединены к объектам, содержащимся в хранилище данных.

5 В частных вариантах предложенной системы хранилище данных может представлять собой базу данных и может содержать вертикальные и горизонтальные таблицы. Вход в вертикальные таблицы может осуществляться с использованием выходных данных механизма анализа одного из вышеупомянутых уровней, а вход в горизонтальные таблицы - с использованием идентификации объекта. Выходными данными могут быть ключи, представляющие соответствующие характеристики объекта, которому эти
10 ключи сопоставлены. Как альтернативный вариант, хранилищем данных может быть, например, файловая система.

При необходимости в системе может также использоваться индекатор, связанный с хранилищем данных, а также быстросействующая кэш-память на полупроводниках и процессор обработки запросов для выполнения запросов по меньшей мере от одного
15 механизма анализа. Кроме того, с механизмами анализа могут быть связаны очереди работ.

В частном предпочтительном варианте индекатор содержит индексы ключей и значения ключей, имеющихся в структурах хранения данных, например, таблицах. Он также может содержать булевы индексы, хранящие значения "да" или "нет" на запросы
20 по форме "имеет ли ключ k значение v?". Кроме того, индекатор может содержать интервальные индексы, хранящие интервалы значений ключей, а также индексы текста. При необходимости индекатор может представлять собой обобщенное воплощение текстового индекатора в виде инвертированного файла, индексирующего Web-документы и обеспечивающего интерфейс прикладного программирования (API) для
25 поиска документов по ключевым словам.

В предпочтительном варианте индекатор может содержать определенные ключи, позволяющие осуществлять запросы в отношении конкретного объекта с применением булевой логики. Кроме того, в индексаторе могут содержаться графические данные, поддерживающие входящие и исходящие запросы.

30 Целесообразно, чтобы снабжение индексов метками и индексирование в индексаторе осуществлялось отдельно.

Недостатками этой системы является отсутствие возможности подключения классификаторов по заданной предметной области, отсутствие блока управления базой данных документов администратором, отсутствие интерфейса для подключения системы
35 поддержки принятия решений по заданной предметной области.

Наиболее близким техническим решением, выбранным в качестве прототипа, является система организации и функционирования базы данных нормативной документации (патент RU 2386166 от 04.02.08 г., МПК G06F 17/30 (2006.1)), содержащая хранилище данных, систему управления базой данных (СУБД), включающую синтаксический
40 анализатор, лингвистический анализатор, XML редактор и интеллектуальный анализатор, предназначенный для проверки по критериям полноты информации, интерфейс анализа и формирования данных, связанный с системным администратором и предназначенный для предварительного анализа документа и объема документации посредством синтаксического, лингвистического и интеллектуального анализаторов
45 для определения характеристических признаков и их составляющих с присвоением кодовых обозначений и введением указанных данных в СУБД для присвоения трем осям трехмерного информационного пространства формируемой базы данных названий характеристических признаков, а также составления соответствующих кластеров с

формированием их кодовых обозначений.

Интерфейс формирования базы данных предназначен для связи системного администратора с редактором XML для перевода документа в XML формат с кодовыми обозначениями, при этом СУБД предназначен для расположения отформатированного документа в базе данных, синтаксический и лингвистический анализаторы предназначены также для определения принадлежности анализируемого документа к каждому кластеру трехмерного информационного пространства, при определении однозначного соответствия документа одному кластеру СУБД предназначен для формирования полного идентификационного номера документа из кода ортов и идентификационного номера и помещения документа в соответствующий кластер трехмерного информационного пространства базы данных, интерфейс пользователя предназначен для доступа к документу или информации с использованием сформированных кодовых обозначений через кластер трехмерного информационного пространства базы данных.

Интеллектуальная надстройка СУБД содержит механизм проверки ссылок, механизм проверки непротиворечивости информации, интеллектуальный анализатор, механизм генерации ссылок.

База данных содержит базу данных оригиналов документов, базу данных документов в формате XML с атрибутами, ключами и сгенерированными ссылками, трехмерное информационное пространство, состоящее из кластеров, образованных составляющими характеристических признаков какой-либо области деятельности, набор синтаксических правил для написания текстовой, нормативной документации, общую терминологическую базу, терминологические базы по предметным областям, критерии полноты документации в кластере, таблицы принадлежности документов к характеристическим признакам, базы шаблонов документов, критерии принадлежности документов к кластерам.

Графический интерфейс содержит интерфейс формирования базы данных, интерфейс анализа базы данных, интерфейс пользователя для поиска и анализа информации и интерфейс редактора для создания документов.

Недостатком этой системы является ограничение пространства классификации документов тремя возможными осями, что оказывается недостаточным для n -мерных (n больше 3) пространств атрибутов нормативно-технических документов. Также отсутствует возможность подключения системы поддержки принятия решений на основе информации, содержащейся в нормативно-технических документах.

Таким образом, возникает задача создания системы структурированного хранения нормативно-технической документации в базе данных, с возможностью многомерной классификации документов, позволяющей осуществлять принятие решений на основе информации, содержащейся в нормативно-технических документах.

Технический результат: повышение упорядоченности хранения нормативно-технических документов, представленных в электронно-цифровой форме, увеличение быстродействия и точности поиска документов; обеспечение возможностей извлечения фактов, описывающих техногенные объекты, формирования новых знаний о техногенных объектах на основе имеющихся, принятия решений на основе анализа нормативной информации; упрощение и оптимизация доступа потребителей к нормативно-технической информации.

Он достигается тем, что система состоит из блока интерфейса администратора, выход которого связан со входом блока ввода документов в базу данных, первый выход которого связан со входом блока хранения полных текстов документов, второй выход

соединен со входом блока хранения фактографической информации о документах, выход блока хранения полных текстов документов связан со входом блока первичной обработки полных текстов документов, первый выход которого соединен со входом блока хранения кодов и представлений терминов (понятий, концептов), второй выход
 5 соединен со входом блока хранения позиций терминов, третий выход соединен со входом блока хранения частот терминов, выходы блоков хранения кодов и представлений терминов (понятий, концептов), хранения позиций терминов, хранения частот терминов связаны с тремя входами блока интерфейса информационно-поисковой системы. Первый выход блока хранения фактографической информации о документах
 10 соединен с четвертым входом блока интерфейса информационно-поисковой системы. Второй выход связан с первым входом блока многомерной классификации документов, остальные входы которого со второго по $n+1$ вход связаны с выходами n классификаторов по соответствующим измерениям многомерного классификационного пространства документов, образующих блок классификаторов. Выход блока
 15 многомерной классификации документов соединен с первым входом блока интерфейса системы поддержки принятия решений. Выход блока интерфейса информационно-поисковой системы связан со вторым входом блока интерфейса системы поддержки принятия решений.

Предлагаемая система изображена на фигуре.

20 Система структурированного хранения нормативно-технической документации в базе данных состоит из блока интерфейса администратора 1, выход которого связан со входом блока ввода документов в базу данных 2, первый выход которого связан со входом блока хранения полных текстов документов 3, второй выход соединен со входом блока хранения фактографической информации о документах 9, выход блока хранения
 25 полных текстов документов 3 связан со входом блока первичной обработки полных текстов документов 4, первый выход которого соединен со входом блока хранения кодов и представлений терминов (понятий, концептов) 5, второй выход соединен со входом блока хранения позиций терминов 6, третий выход соединен со входом блока хранения частот терминов 7, выходы блоков хранения кодов и представлений терминов
 30 (понятий, концептов) 5, хранения позиций терминов 6, хранения частот терминов 7 связаны с тремя входами блока интерфейса информационно-поисковой системы 8. Первый выход блока хранения фактографической информации о документах 9 соединен с четвертым входом блока интерфейса информационно-поисковой системы 8. Второй выход связан с первым входом блока многомерной классификации документов 10,
 35 остальные входы которого со второго по $n+1$ вход связаны с выходами n классификаторов 11.1, 11.2, ..., 11.1, ..., 11.n по соответствующим измерениям многомерного классификационного пространства документов, образующих блок классификаторов 11. Блок многомерной классификации 10 реализован в системе для повышения точности поиска и релевантности поисковой выдачи запросу. Выход блока
 40 многомерной классификации документов 10 соединен с первым входом блока интерфейса системы поддержки принятия решений 12. Выход блока интерфейса информационно-поисковой системы 8 связан со вторым входом блока интерфейса системы поддержки принятия решений 12.

Система структурированного хранения нормативно-технической документации в
 45 базе данных работает следующим образом. С помощью блока интерфейса администратора 1 администратор системы вводит нормативно-технические документы (НТД) и сопутствующую им фактографическую информацию в блок ввода документов в базу данных 2. Также возможен пакетный ввод документов, загружаемых из внешних

источников. Блок ввода документов в базу данных 2 осуществляет передачу информации на хранение блокам хранения фактографической информации о документах 9 и хранения полных текстов документов 3, при этом блок хранения фактографической информации о документах 9 обеспечивает хранение только атрибутов документов, а блок хранения полных текстов документов 3 - только полных текстов документов в электронно-цифровой форме в различных текстовых и графических форматах представления документальной информации.

Далее необработанные данные, содержащиеся в полных текстах нормативно-технических документов, передаются в блок первичной обработки полных текстов документов 4, который осуществляет первоначальное извлечение текста документов, представленных в различных текстовых и графических форматах, формирует последовательности терминов, обрабатывает их и передает обработанные данные блокам хранения кодов и представлений терминов (понятий, концептов) 5, хранения позиций терминов 6, хранения частот терминов 7. Блок хранения кодов и представлений терминов (понятий, концептов) 5 накапливает символьные представления терминов, понятий, концептов заданной предметной области, формирует уникальный код термина (понятия, концепта), устанавливает и запоминает отношения между терминами, включая, но не ограничиваясь, отношения синонимичности, отношения «часть-целое», «частное-общее», «сокращение-сокращаемое», отношения между равными по значению словами и словосочетаниями и др. Блок также формирует и наполняет классы тезауруса. Блок хранения позиций терминов 6 накапливает информацию о взаимном расположении терминов в последовательностях, образующих документ, с учетом возможного составного характера символьных представлений терминов. Блок хранения частот терминов 7 накапливает статистическую информацию о последовательностях терминов, включая, но не ограничиваясь, частоты терминов в пределах одного документа, частоты терминов в пределах совокупности документов, частоты терминов во всем фонде документов, частоты документов, содержащих определенный термин, с учетом возможного составного характера символьных представлений терминов. Совместно блоки 5, 6 и 7 формируют необходимый объем информации для использования в подключаемых информационно-поисковых системах посредством блока интерфейса информационно-поисковой системы 8. Блок интерфейса информационно-поисковой системы 8 обеспечивает доступ дополнительно подключаемых информационно-поисковых систем к структурированной информации нормативно-технических документов. Кроме того, он формирует выдачу информации по запросу блока интерфейса системы принятия решений 12. Блок хранения фактографической информации о документах 9 передает значения атрибутов документов, образующих многомерное пространство атрибутов, блоку многомерной классификации 10. Блок многомерной классификации 10 осуществляет присвоение классификационных признаков документам на основании соответствия значений их атрибутов значениям классификационных рубрик, которые поступают из блока классификаторов 11.

При этом для каждой оси многомерного пространства классификации используется отдельный классификатор 11.1, 11.2, ..., 11.i, ..., 11.n, включая, но не ограничивая, классификаторы видов документов, типов документов, предметных областей (областей применения), категорий документов, тематических направлений, техногенных объектов, с которыми связаны документы, ситуаций контроля, отраслей, организаций, утвердивших документ и др. Для классификации документов по многомерному тематическому подпространству блок многомерной классификации 10 создает обучающую выборку, для которой отбираются несколько наиболее представительных документов для каждого

класса таким образом, чтобы в совокупности они достаточно полно описывали предметную область класса. Полученная выборка документов анализируется с целью выделения наиболее значимых слов, приведения их к словарным формам, формирования распределений частот (весов) полученных ключевых слов по документам и классам, после чего для каждого класса составляется характеристика (профиль). Отнесение документов к тому или иному классу осуществляется на основании расчета степени подобия документа профилю класса. В качестве меры сходства документа и профиля класса используется коэффициент косинуса (имеется в виду косинус угла между векторами документа и профиля в многомерном пространстве терминов).

Результаты классификации документов из блока многомерной классификации 10 и результаты обработки полных текстов документов из блока интерфейса информационно-поисковой системы 8 поступают в блок интерфейса системы поддержки принятия решений 12, который осуществляет формирование агрегированной информации для последующего принятия решений в подключаемой системе поддержки принятия решений (СППР). Таким образом, предлагаемая система предоставляет для СППР функциональные возможности базы знаний. Блок интерфейса системы поддержки принятия решений 12 обеспечивает хранение отношений, определяемых конкретной областью технического регулирования, включая, но не ограничиваясь, отношения «объект контроля - область контроля», «контрольный вопрос - объект контроля», «выдержка из НТД - контрольный вопрос» и др. Обследование объекта контроля с помощью предлагаемой системы осуществляется путем получения ответов на определенный набор контрольных вопросов. Выбор вопросов определяется характером объекта контроля. Каждый вопрос связан с положениями нормативно-технических документов, которые могут быть представлены специалисту по запросу с помощью блока интерфейса информационно-поисковой системы 8. В блоке интерфейса системы поддержки принятия решений 12 формируется отчет, в котором информация о нарушениях требований сопровождается ссылками на соответствующие НТД.

Таким образом, предлагаемая система структурированного хранения нормативно-технической документации в базе данных, обладая функциональными возможностями многомерной классификации документов и агрегирования информации для обеспечения принятия решений на основе информации, содержащейся в нормативно-технических документах, обеспечивает повышение упорядоченности хранения нормативно-технических документов, представленных в электронно-цифровой форме, увеличение быстродействия и точности поиска документов; обеспечение возможностей извлечения фактов, описывающих техногенные объекты, формирования новых знаний о техногенных объектах на основе имеющихся, принятия решений на основе анализа нормативной информации; упрощение и оптимизация доступа потребителей к нормативно-технической информации.

Система структурированного хранения нормативно-технической документации в базе данных используется на предприятии.

(57) Реферат

Система структурированного хранения нормативно-технической документации в базе данных предназначена для хранения, поиска и анализа нормативно-технических документов, представленных в электронно-цифровой форме. Система содержит блоки интерфейса администратора (1), ввода документов в базу данных (2), хранения полных текстов документов (3), хранения фактографической информации о документах (9), первичной обработки полных текстов документов (4), хранения кодов и представлений

терминов (5), хранения позиций терминов (6), хранения частот терминов (7), интерфейса информационно-поисковой системы (8), интерфейса системы поддержки принятия решений (12), многомерной классификации документов (10) с произвольным числом классификационных измерений n больше 3, классификаторов (11) с количеством классификаторов многомерного классификационного пространства документов n больше 3. Входы блока многомерной классификации документов (10) со второго по n плюс 1 вход связаны с выходами n классификаторов блока классификаторов (11). Часть классификаторов сформирована автоматически путем анализа содержания обучающей выборки, формирования профилей классов обучающей выборки, анализа содержания полных текстов хранимых документов, автоматического сопоставления документов профилю класса на основе меры близости для их тематической классификации. Выход блока многомерной классификации документов (10) и выход блока интерфейса информационно-поисковой системы (8) связаны с входами блока интерфейса системы поддержки принятия решений (12), содержащего модули сопоставления объектов контроля и области контроля, сопоставления контрольных вопросов и объекта контроля, сопоставления фрагментов нормативно-технических документов и контрольного вопроса. Система обеспечивает повышение упорядоченности хранения документов, увеличение быстродействия и точности поиска документов, возможность извлечения фактов, которые описывают техногенные объекты и формируют новые знания о техногенных объектах на основе имеющихся, принятие решений на основе анализа нормативной информации, упрощение и оптимизацию доступа потребителей к нормативно-технической информации. 1 ил.

25

30

35

40

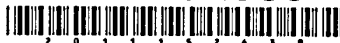
45

РЕФЕРАТ

(57) Система структурированного хранения нормативно-технической документации в базе данных предназначена для хранения, поиска и анализа нормативно-технических документов, представленных в электронно-цифровой форме. Система содержит блоки интерфейса администратора (1), ввода документов в базу данных (2), хранения полных текстов документов (3), хранения фактографической информации о документах (9), первичной обработки полных текстов документов (4), хранения кодов и представлений терминов (5), хранения позиций терминов (6), хранения частот терминов (7), интерфейса информационно-поисковой системы (8), интерфейса системы поддержки принятия решений (12), многомерной классификации документов (10) с произвольным числом классификационных измерений n больше 3, классификаторов (11) с количеством классификаторов многомерного классификационного пространства документов n больше 3. Входы блока многомерной классификации документов (10) со второго по n плюс 1 вход связаны с выходами n классификаторов блока классификаторов (11). Часть классификаторов сформирована автоматически путем анализа содержания обучающей выборки, формирования профилей классов обучающей выборки, анализа содержания полных текстов хранимых документов, автоматического сопоставления документов профилю класса на основе меры близости для их тематической классификации. Выход блока многомерной классификации документов (10) и выход блока интерфейса информационно-поисковой системы (8) связаны с входами блока интерфейса системы поддержки принятия решений (12), содержащего модули сопоставления объектов контроля и области контроля, сопоставления контрольных вопросов и объекта контроля, сопоставления фрагментов нормативно-технических документов и контрольного вопроса. Система обеспечивает повышение упорядоченности хранения документов, увеличение быстродействия и точности поиска документов, возможность извлечения фактов, которые описывают техногенные объекты и формируют новые знания о техногенных объектах на основе имеющихся, принятие решений на основе анализа нормативной информации, упрощение и оптимизацию доступа потребителей к нормативно-технической информации. 1 ил.

Референт: Генеральный директор
ООО «Газпром газнадзор»
В.Н. Медведев





СИСТЕМА СТРУКТУРИРОВАННОГО ХРАНЕНИЯ НОРМАТИВНО-ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ В БАЗЕ ДАННЫХ

Полезная модель относится к вычислительной технике и может быть использована в промышленности, на транспорте, на предприятиях связи и в других отраслях экономики, охватываемых техническим регулированием, для повышения упорядоченности хранения нормативно-технических документов, представленных в электронно-цифровой форме, увеличения быстродействия и точности поиска документов; обеспечения возможностей извлечения фактов, описывающих техногенные объекты, формирования новых знаний о техногенных объектах на основе имеющихся, принятия решений на основе анализа нормативной информации; упрощения и оптимизации доступа потребителей к нормативно-технической информации.

Известна справочная правовая система хранения и поиска данных (патент RU 2223537 от 22.11.01 г., МПК⁷ G06F 17/30, 17/40), содержащая блок выбора вида поиска, формирователь запроса, один выход которого связан с блоком проведения поиска, вход-выход которого соединен через соответствующие шины с базами данных системы, блоком отображения и контроллером, предназначенным для управления поиском данных, в систему введены блок выбора условий поиска, вход которого связан с выходом блока выбора вида поиска, первый выход которого связан со входом формирователя запроса, второй выход соединен с входом формирователя атрибутов фильтра, первый вход-выход которого связан со вторым входом-выходом формирователя запроса, второй вход-выход - с блоком памяти атрибутов фильтра, первый вход-выход формирователя запроса соединен с блоком памяти запросов.

Базами данных системы являются база данных правовой информации и/или база данных экономической информации, база данных энциклопедии ситуаций и/или база данных толкового словаря, и/или база данных последних изменений в законодательстве.

При этом блок выбора вида поиска выполнен с возможностью навигации по базам данных с поэтапным представлением отклассифицированных списков запрашиваемых документов и/или фрагментов документов. Система также может включать средство сортировки документов, связанное с памятью, указанное средство и память подключены через соответствующие шины системы к соответствующим блокам системы.

Средство сортировки документов предназначено для сортировки документов по дате издания документов или по юридической силе.

Система включает также средство поиска документов по заданному числу сходных признаков, связанное по соответствующим шинам с соответствующими блоками системы.

Недостатком этой системы является то, что она не обеспечивает построение индекса массива документов, что снижает быстродействие системы, а также не гарантирует полноту и точность поисковой выдачи.

Известна также вычислительная система для интеллектуального анализа данных – хранилище данных для основанной на знаниях системы извлечения информации из данных (патент RU 2297665 от 28.04.03 г., МПК⁸ G06F 17/30), включающая в себя, по меньшей мере, одно хранилище данных, механизмы анализа более низкого уровня, механизмы анализа более высокого уровня, индексатор.

Система извлечения информации из данных, включающая в себя: по меньшей мере, одно хранилище, или склад данных, в котором содержатся объекты; по меньшей мере, один механизм анализа более низкого уровня, связанный с хранилищем данных и генерирующий выходные данные на основе первого набора правил, реализованных в указанном механизме анализа более низкого уровня; и, по меньшей мере, один механизм анализа более высокого уровня, принимающий выходные данные механизма анализа более низкого уровня и генерирующий свои выходные данные на основе второго набора правил, реализованных в указанном механизме анализа более высокого уровня, причем выходные данные механизмов анализа более низкого и более высокого уровней присоединены к объектам, содержащимся в хранилище данных.

В частных вариантах предложенной системы хранилище данных может представлять собой базу данных и может содержать вертикальные и горизонтальные таблицы. Вход в вертикальные таблицы может осуществляться с использованием выходных данных механизма анализа одного из вышеупомянутых уровней, а вход в горизонтальные таблицы - с использованием идентификации объекта. Выходными данными могут быть ключи, представляющие соответствующие характеристики объекта, которому эти ключи сопоставлены. Как альтернативный вариант, хранилищем данных может быть, например, файловая система.

При необходимости в системе может также использоваться индексатор, связанный с хранилищем данных, а также быстродействующая кэш-память на полупроводниках и процессор обработки запросов для выполнения запросов по меньшей мере от одного механизма анализа. Кроме того, с механизмами анализа могут быть связаны очереди работ.

В частном предпочтительном варианте индексатор содержит индексы ключей и значения ключей, имеющих в структурах хранения данных, например, таблицах. Он также может содержать булевы индексы, хранящие значения "да" или "нет" на запросы по форме "имеет ли ключ k значение v?". Кроме того, индексатор может содержать интервальные индексы, хранящие интервалы значений ключей, а также индексы текста. При необходимости индексатор может представлять собой обобщенное воплощение текстового индексатора в виде инвертированного файла, индексирующего Web-документы и обеспечивающего интерфейс прикладного программирования (API) для поиска документов по ключевым словам.

В предпочтительном варианте индексатор может содержать определенные ключи, позволяющие осуществлять запросы в отношении конкретного объекта с применением булевой логики. Кроме того, в индексаторе могут содержаться графические данные, поддерживающие входящие и исходящие запросы.

Целесообразно, чтобы снабжение индексов метками и индексирование в индексаторе осуществлялось отдельно.

Недостатками этой системы является отсутствие возможности подключения классификаторов по заданной предметной области, отсутствие блока управления базой данных документов администратором, отсутствие интерфейса для подключения системы поддержки принятия решений по заданной предметной области.

Наиболее близким техническим решением, выбранным в качестве прототипа, является система организации и функционирования базы данных нормативной документации (патент RU 2386166 от 04.02.08 г., МПК G06F 17/30 (2006.1)), содержащая хранилище данных, систему управления базой данных (СУБД), включающую синтаксический анализатор, лингвистический анализатор, XML редактор и интеллектуальный анализатор, предназначенный для проверки по критериям полноты информации, интерфейс анализа и формирования данных, связанный с системным администратором и предназначенный для предварительного анализа документа и объема документации посредством синтаксического, лингвистического и интеллектуального анализаторов для определения характеристических признаков и их составляющих с присвоением кодовых обозначений и введением указанных данных в СУБД для присвоения трем осям трехмерного информационного пространства формируемой базы данных названий характеристических признаков, а также составления соответствующих кластеров с формированием их кодовых обозначений.

Интерфейс формирования базы данных предназначен для связи системного администратора с редактором XML для перевода документа в XML формат с кодовыми обозначениями, при этом СУБД предназначен для расположения отформатированного документа в базе данных, синтаксический и лингвистический анализаторы предназначены также для определения принадлежности анализируемого документа к каждому кластеру трехмерного информационного пространства, при определении однозначного соответствия документа одному кластеру СУБД предназначен для формирования полного идентификационного номера документа из кода ортов и идентификационного номера и помещения документа в соответствующий кластер трехмерного информационного пространства базы данных, интерфейс пользователя предназначен для доступа к документу или информации с использованием сформированных кодовых обозначений через кластер трехмерного информационного пространства базы данных.

Интеллектуальная надстройка СУБД содержит механизм проверки ссылок, механизм проверки непротиворечивости информации, интеллектуальный анализатор, механизм генерации ссылок.

База данных содержит базу данных оригиналов документов, базу данных документов в формате XML с атрибутами, ключами и сгенерированными ссылками, трехмерное информационное пространство, состоящее из кластеров, образованных составляющими характеристических признаков какой-либо области деятельности, набор синтаксических правил для написания текстовой, нормативной документации, общую терминологическую базу, терминологические базы по предметным областям, критерии полноты документации в кластере, таблицы принадлежности документов к характеристическим признакам, базы шаблонов документов, критерии принадлежности документов к кластерам.

Графический интерфейс содержит интерфейс формирования базы данных, интерфейс анализа базы данных, интерфейс пользователя для поиска и анализа информации и интерфейс редактора для создания документов.

Недостатком этой системы является ограничение пространства классификации документов тремя возможными осями, что оказывается недостаточным для n -мерных (n больше 3) пространств атрибутов нормативно-технических документов. Также отсутствует возможность подключения системы поддержки принятия решений на основе информации, содержащейся в нормативно-технических документах.

Таким образом, возникает задача создания системы структурированного хранения нормативно-технической документации в базе данных, с возможностью многомерной классификации документов, позволяющей осуществлять принятие решений на основе информации, содержащейся в нормативно-технических документах.

Технический результат: повышение упорядоченности хранения нормативно-технических документов, представленных в электронно-цифровой форме, увеличение быстродействия и точности поиска документов; обеспечение возможностей извлечения фактов, описывающих техногенные объекты, формирования новых знаний о техногенных объектах на основе имеющихся, принятия решений на основе анализа нормативной информации; упрощение и оптимизация доступа потребителей к нормативно-технической информации.

Он достигается тем, что система состоит из блока интерфейса администратора, выход которого связан со входом блока ввода документов в базу данных, первый выход которого связан со входом блока хранения полных текстов документов, второй выход соединен со входом блока хранения фактографической информации о документах, выход блока хранения полных текстов документов связан со входом блока первичной обработки полных текстов документов, первый выход которого соединен со входом блока хранения кодов и представлений терминов (понятий, концептов), второй выход соединен со входом блока хранения позиций терминов, третий выход соединен со входом блока хранения частот терминов, выходы блоков хранения кодов и представлений терминов (понятий, концептов), хранения позиций терминов, хранения частот терминов связаны с тремя входами блока интерфейса информационно-поисковой системы. Первый выход блока хранения фактографической информации о документах соединен с четвертым входом блока интерфейса информационно-поисковой системы. Второй выход связан с первым входом блока многомерной классификации документов, остальные входы которого со второго по $n+1$ вход связаны с выходами n классификаторов по соответствующим измерениям многомерного классификационного пространства документов, образующих блок классификаторов. Выход блока многомерной классификации документов соединен с первым входом блока интерфейса системы поддержки принятия решений. Выход блока интерфейса информационно-поисковой системы связан со вторым входом блока интерфейса системы поддержки принятия решений.

Предлагаемая система изображена на фигуре.

Система структурированного хранения нормативно-технической документации в базе данных состоит из блока интерфейса администратора 1, выход которого связан со входом блока ввода документов в базу данных 2, первый выход которого связан со входом блока хранения полных текстов документов 3, второй выход соединен со входом блока хранения фактографической информации о документах 9, выход блока хранения полных текстов документов 3 связан со входом блока первичной обработки полных текстов документов 4, первый выход которого соединен со входом блока хранения кодов и представлений терминов (понятий, концептов) 5, второй выход соединен со входом блока хранения позиций терминов 6, третий выход соединен со входом блока хранения частот терминов 7, выходы блоков хранения кодов и представлений терминов (понятий, концептов) 5, хранения позиций терминов 6, хранения частот терминов 7 связаны с тремя входами блока интерфейса информационно-поисковой системы 8. Первый выход блока хранения фактографической информации о документах 9 соединен с четвертым входом блока интерфейса информационно-поисковой системы 8. Вторым выходом связан с первым входом блока многомерной классификации документов 10, остальные входы которого со второго по $n+1$ вход связаны с выходами n классификаторов 11.1, 11.2, ..., 11.i, ..., 11.n по соответствующим измерениям многомерного классификационного пространства документов, образующих блок классификаторов 11. Блок многомерной классификации 10 реализован в системе для повышения точности поиска и релевантности поисковой выдачи запросу. Выход блока многомерной классификации документов 10 соединен с первым входом блока интерфейса системы поддержки принятия решений 12. Выход блока интерфейса информационно-поисковой системы 8 связан со вторым входом блока интерфейса системы поддержки принятия решений 12.

Система структурированного хранения нормативно-технической документации в базе данных работает следующим образом. С помощью блока интерфейса администратора 1 администратор системы вводит нормативно-технические документы (НТД) и сопутствующую им фактографическую информацию в блок ввода документов в базу данных 2. Также возможен пакетный ввод документов, загружаемых из внешних источников. Блок ввода документов в базу данных 2 осуществляет передачу информации на хранение блокам хранения фактографической информации о документах 9 и хранения полных текстов документов 3, при этом блок хранения фактографической информации о документах 9 обеспечивает хранение только атрибутов документов, а блок хранения полных текстов документов 3 – только полных текстов документов в электронно-цифровой форме в различных текстовых и графических форматах представления документальной информации.

Далее необработанные данные, содержащиеся в полных текстах нормативно-технических документов, передаются в блок первичной обработки полных текстов документов 4, который осуществляет первоначальное извлечение текста документов, представленных в различных текстовых и графических форматах, формирует последовательности терминов, обрабатывает их и передает обработанные данные блокам хранения кодов и представлений терминов (понятий, концептов) 5, хранения позиций терминов 6, хранения частот терминов 7. Блок хранения кодов и представлений терминов (понятий, концептов) 5 накапливает символьные представления терминов, понятий, концептов заданной предметной области, формирует уникальный код термина (понятия, концепта), устанавливает и запоминает отношения между терминами, включая, но не ограничиваясь, отношения синонимичности, отношения «часть-целое», «частное-общее», «сокращение-сокращаемое», отношения между равными по значению словами и словосочетаниями и др. Блок также формирует и наполняет классы тезауруса. Блок хранения позиций терминов 6 накапливает информацию о взаимном расположении терминов в последовательностях, образующих документ, с учетом возможного составного характера символьных представлений терминов. Блок хранения частот терминов 7 накапливает статистическую информацию о последовательностях терминов, включая, но не ограничиваясь, частоты терминов в пределах одного документа, частоты терминов в пределах совокупности документов, частоты терминов во всем фонде документов, частоты документов, содержащих определенный термин, с учетом возможного составного характера символьных представлений терминов. Совместно блоки 5,6 и 7 формируют необходимый объем информации для использования в подключаемых информационно-поисковых системах посредством блока интерфейса информационно-поисковой системы 8. Блок интерфейса информационно-поисковой системы 8 обеспечивает доступ дополнительно подключаемых информационно-поисковых систем к структурированной информации нормативно-технических документов. Кроме того, он формирует выдачу информации по запросу блока интерфейса системы принятия решений 12. Блок хранения фактографической информации о документах 9 передает значения атрибутов документов, образующих многомерное пространство атрибутов, блоку многомерной классификации 10. Блок многомерной классификации 10 осуществляет присвоение классификационных признаков документам на основании соответствия значений их атрибутов значениям классификационных рубрик, которые поступают из блока классификаторов 11.

При этом для каждой оси многомерного пространства классификации используется отдельный классификатор 11.1, 11.2, ..., 11.i, ..., 11.n, включая, но не ограничивая, классификаторы видов документов, типов документов, предметных областей (областей применения), категорий документов, тематических направлений, техногенных объектов, с которыми связаны документы, ситуаций контроля, отраслей, организаций, утвердивших документ и др. Для классификации документов по многомерному тематическому подпространству блок многомерной классификации 10 создает обучающую выборку, для которой отбираются несколько наиболее представительных документов для каждого класса таким образом, чтобы в совокупности они достаточно полно описывали предметную область класса. Полученная выборка документов анализируется с целью выделения наиболее значимых слов, приведения их к словарным формам, формирования распределений частот (весов) полученных ключевых слов по документам и классам, после чего для каждого класса составляется характеристика (профиль). Отнесение документов к тому или иному классу осуществляется на основании расчета степени подобия документа профилю класса. В качестве меры сходства документа и профиля класса используется коэффициент косинуса (имеется в виду косинус угла между векторами документа и профиля в многомерном пространстве терминов).

Результаты классификации документов из блока многомерной классификации 10 и результаты обработки полных текстов документов из блока интерфейса информационно-поисковой системы 8 поступают в блок интерфейса системы поддержки принятия решений 12, который осуществляет формирование агрегированной информации для последующего принятия решений в подключаемой системе поддержки принятия решений (СППР). Таким образом, предлагаемая система предоставляет для СППР функциональные возможности базы знаний. Блок интерфейса системы поддержки принятия решений 12 обеспечивает хранение отношений, определяемых конкретной областью технического регулирования, включая, но не ограничиваясь, отношения «объект контроля – область контроля», «контрольный вопрос – объект контроля», «выдержка из НТД – контрольный вопрос» и др. Обследование объекта контроля с помощью предлагаемой системы осуществляется путем получения ответов на определенный набор контрольных вопросов. Выбор вопросов определяется характером объекта контроля. Каждый вопрос связан с положениями нормативно-технических документов, которые могут быть представлены специалисту по запросу с помощью блока интерфейса информационно-поисковой системы 8. В блоке интерфейса системы поддержки принятия решений 12 формируется отчет, в котором информация о нарушениях требований сопровождается ссылками на соответствующие НТД.

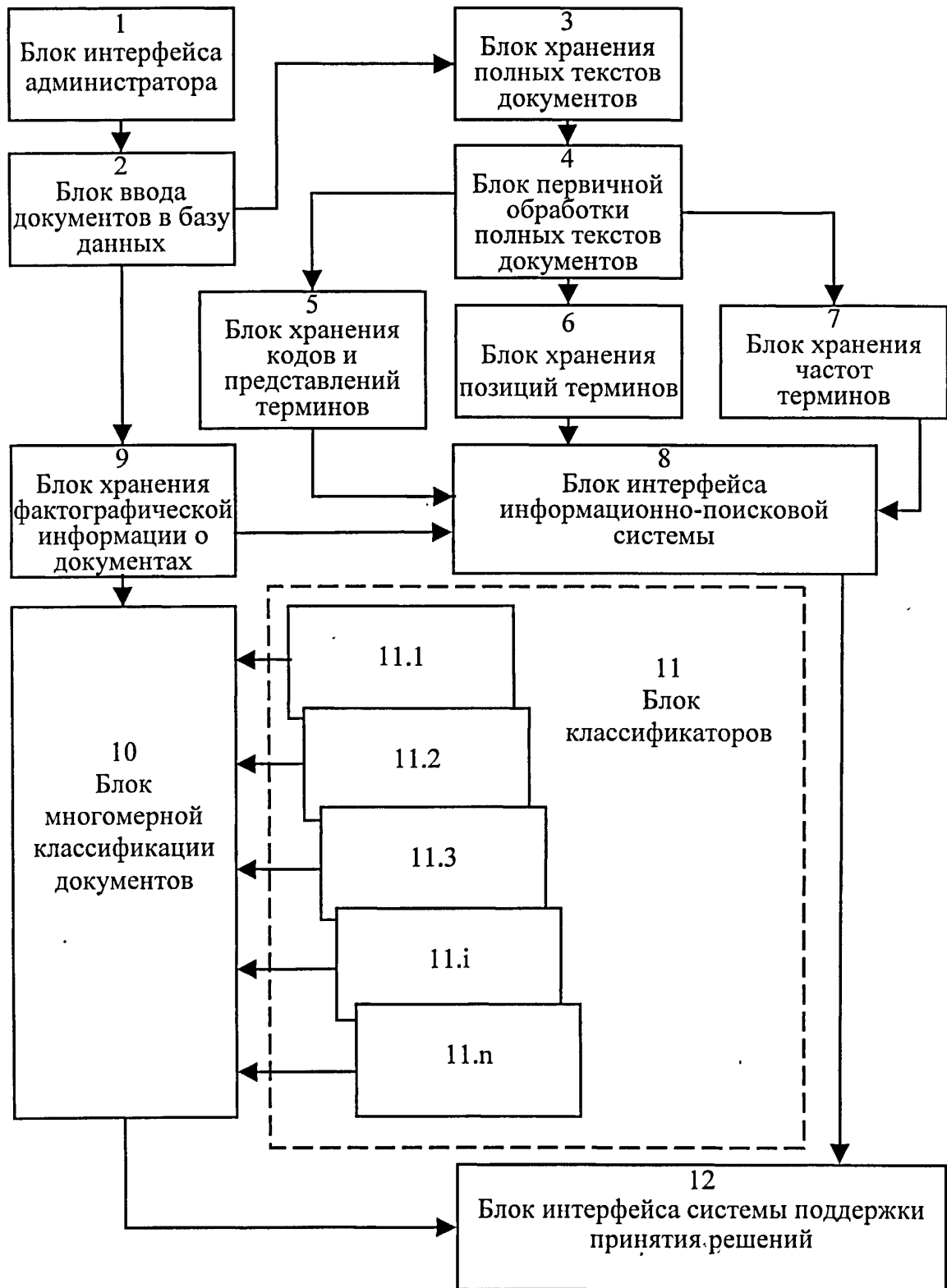
Таким образом, предлагаемая система структурированного хранения нормативно-технической документации в базе данных, обладая функциональными возможностями многомерной классификации документов и агрегирования информации для обеспечения принятия решений на основе информации, содержащейся в нормативно-технических документах, обеспечивает повышение упорядоченности хранения нормативно-технических документов, представленных в электронно-цифровой форме, увеличение быстродействия и точности поиска документов; обеспечение возможностей извлечения фактов, описывающих техногенные объекты, формирования новых знаний о техногенных объектах на основе имеющихся, принятия решений на основе анализа нормативной информации; упрощение и оптимизация доступа потребителей к нормативно-технической информации.

Система структурированного хранения нормативно-технической документации в базе данных используется на предприятии.

Заявитель: ООО «Газпром газнадзор»



Система структурированного хранения
нормативно-технической документации
в базе данных



Фигура