



**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: **2009103145/08**, **20.07.2007**

(24) Дата начала отсчета срока действия патента:
20.07.2007

Приоритет(ы):

(30) Конвенционный приоритет:
31.07.2006 US 11/496,650

(43) Дата публикации заявки: **10.08.2010** Бюл. № 22

(45) Опубликовано: **27.05.2012** Бюл. № 15

(56) Список документов, цитированных в отчете о
поиске: **US 2002/0178152 A1**, **28.11.2002. RU**
2236699 C1, **20.09.2004. JP 2000029902 A**,
28.01.2000. KR 20030044949 A, **09.06.2003.**

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: **30.01.2009**

(86) Заявка РСТ:
US 2007/016435 (20.07.2007)

(87) Публикация заявки РСТ:
WO 2008/016491 (07.02.2008)

Адрес для переписки:

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

(72) Автор(ы):

**АЗЗАМ Салиха (US),
ХАМФРИЗ Кевин Уильям (US)**

(73) Патентообладатель(и):

МАЙКРОСОФТ КОРПОРЕЙШН (US)

(54) ОПТИМИЗАЦИЯ ИЗВЛЕЧЕНИЯ ФАКТА С ИСПОЛЬЗОВАНИЕМ МНОГОЭТАПНОГО ПОДХОДА

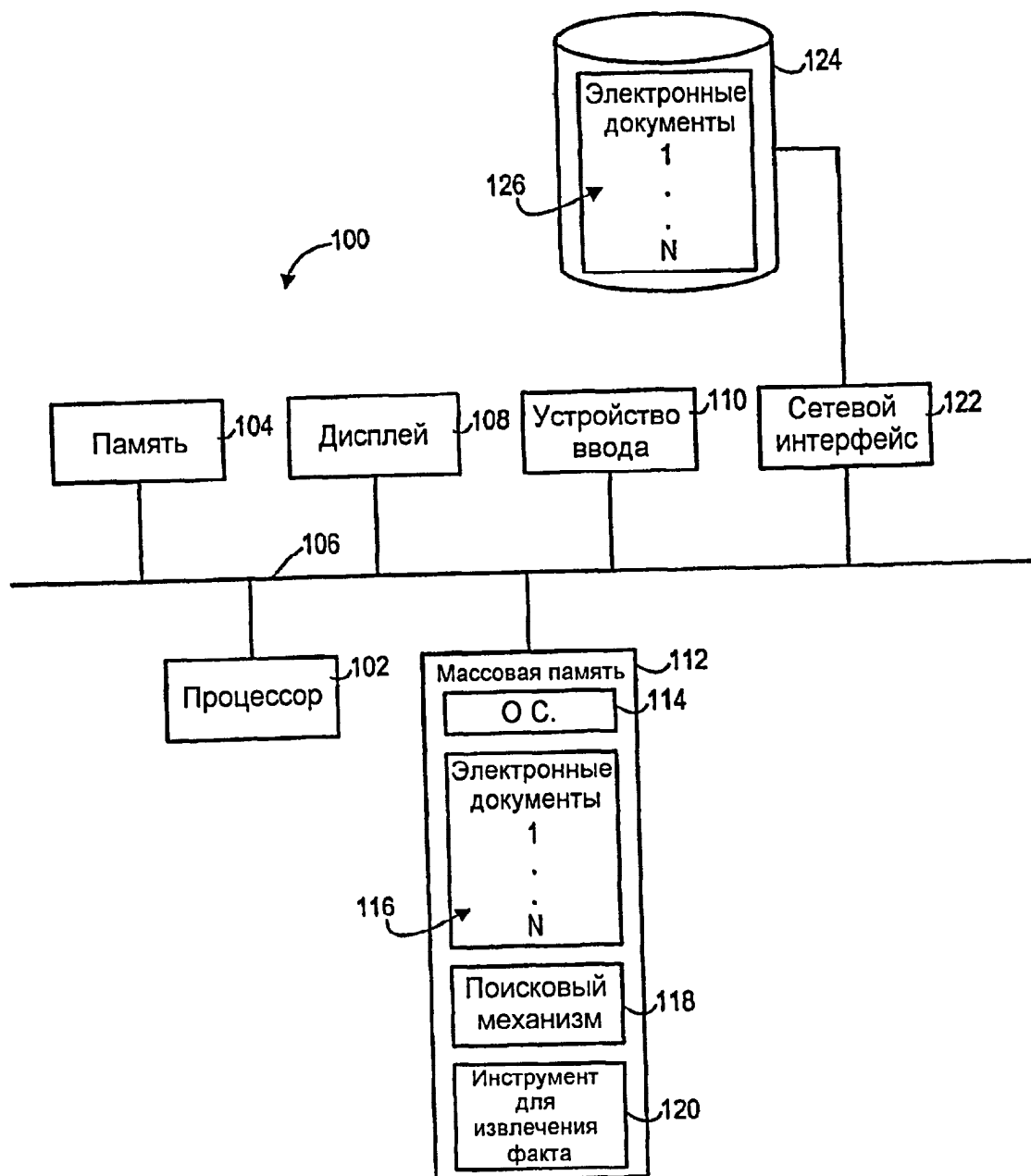
(57) Реферат:

Изобретение относится к способу и устройству для проведения информационного поиска. Техническим результатом является повышение достоверности результатов поиска. Из электронных документов извлекаются факты посредством распознавания фактографических описаний с использованием таблицы слов факта, сопоставляемых со словами электронных документов. Слова этих фактографических описаний могут быть обеспечены признаком соответствующей части речи. Далее выполняется более подробный

анализ этих фактографических описаний, а не всего электронного документа, и, в частности, текста, окружающего соответствующие слова факта. Анализ может включать в себя идентификацию лингвистических элементов каждого словосочетания и определение их роли как подлежащего или как дополнения. Могут применяться правила исключения для удаления тех словосочетаний, которые, скорее всего, не являются частью фактов, причем эти правила исключения частично основаны на лингвистических элементах. К оставшимся словосочетаниям могут быть применены

правила оценки, и для тех словосочетаний, которые имеют оценку, превышающую порог, соответствующая часть предложения, целое предложение, абзац или другая часть

документа могут быть представлены как представляющие один или несколько фактов. 3 н. и 17 з.п. ф-лы, 6 ил., 4 табл.



Фиг.1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.

G06F 17/21 (2006.01)*G06F 17/30* (2006.01)**(12) ABSTRACT OF INVENTION**(21)(22) Application: **2009103145/08, 20.07.2007**(24) Effective date for property rights:
20.07.2007

Priority:

(30) Convention priority:
31.07.2006 US 11/496,650(43) Application published: **10.08.2010 Bull. 22**(45) Date of publication: **27.05.2012 Bull. 15**(85) Commencement of national phase: **30.01.2009**(86) PCT application:
US 2007/016435 (20.07.2007)(87) PCT publication:
WO 2008/016491 (07.02.2008)

Mail address:

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

(72) Inventor(s):

**AZZAM Salikha (US),
KhAMFRIZ Kevin Uill'jam (US)**

(73) Proprietor(s):

MAJKROSOFT KORPOREJShN (US)**(54) OPTIMISATION OF FACT EXTRACTION USING MULTI-STAGE APPROACH**

(57) Abstract:

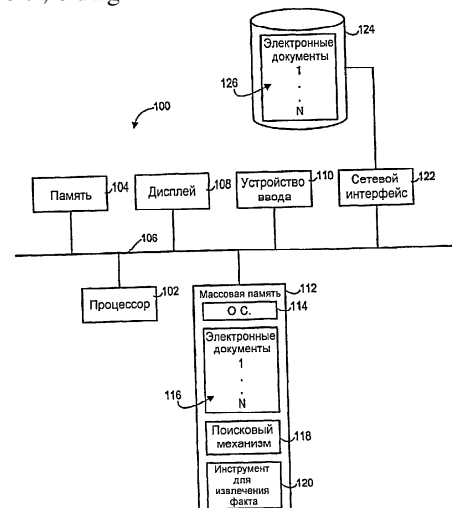
FIELD: information technology.

SUBSTANCE: facts are extracted from electronic documents by recognising factual descriptions using a fact-word table to match to words of the electronic documents. The words of those factual descriptions may be tagged with the appropriate part of speech. More detailed analysis is then performed on those factual descriptions, rather than on the entire electronic document, and particularly to the text in the neighbourhood of the fact-word matches. The analysis may involve identifying the linguistic constituents of each phrase and determining the role as either subject or object. Exclusion rules may be applied to eliminate those phrases unlikely to be part of facts, the exclusion rules being based in part on the linguistic constituents. Scoring rules may be applied to remaining phrases, and for those phrases having a score in excess of a threshold, the corresponding sentence part, whole sentence,

paragraph, or other document portion may be presented as representing one or more facts.

EFFECT: more accurate search results.

20 cl, 6 dwg



Фиг.1

Уровень техники

Электронные документы могут содержать смесь фактов и мнений. Время от времени читателя могут интересовать только факты, или ему может потребоваться идентифицировать факты. Например, пользователю, выполняющему поиск информации в режиме онлайн, может потребоваться получить факты по конкретной теме настолько быстро и эффективно, насколько возможно. Однако представление списка web-страниц или других электронных документов, которые относятся к используемым термам поиска, требует, чтобы пользователь сам исследовал каждую web-страницу или другой электронный документ и отличал факты от мнений или субъективной информации.

Были сделаны попытки извлечения факта. Однако точное извлечение факта может быть медленным и неэффективным даже для быстродействующих компьютеров сервера. При таких попытках извлечения факта, как правило, применяют лингвистический анализ ко всему содержимому электронного документа для извлечения тех фактов, которые он может содержать. При применении извлечения факта к сотням или тысячам электронных документов, количество времени, необходимое для достижения результата, может быть недопустимым.

Сущность изобретения

С использованием многоэтапного подхода варианты осуществления обеспечивают оптимизацию извлечения факта. Бегло просматриваются электронные документы для обнаружения фактографических описаний, которые, скорее всего, содержат факты с использованием таблицы слов факта для сопоставления с термами в предложениях электронных документов для получения набора фактографических описаний. После этого можно выполнить дополнительный анализ, включающий в себя определение лингвистических элементов, например синтаксических элементов и/или семантики, в окружении этого набора фактографических описаний, а не во всем документе. Соответственно, за счет отказа от сложного лексического и синтаксического анализа всего документа для каждого представляющего интерес электронного документа обеспечивается экономия времени.

В этом кратком описании представлен в упрощенной форме перечень понятий, которые также описаны ниже в подробном описании. Это краткое описание не предназначено для определения основных или существенных признаков заявленного объекта изобретения, равно как и для использования в качестве вспомогательного средства в определении объема заявленного объекта изобретения.

Краткое описание чертежей

На фиг.1 изображен пример компьютерной системы для реализации вариантов осуществления.

На фиг.2 изображен пример последовательности операций поиска, включающей в себя представление фактов, которые были извлечены до поиска.

На фиг.3 изображен пример последовательности операций поиска, включающей в себя представление фактов, которые были извлечены во время поиска.

На фиг.4 изображен пример последовательности операций множества этапов извлечения факта.

На фиг.5 изображен пример более подробной последовательности операций множества этапов извлечения факта.

На фиг.6 изображен пример экранного устройства отображения, обеспечивающего результаты поиска, которые включают в себя представление фактов, полученных из электронных документов, обнаруженных при поиске.

Подробное описание

Варианты осуществления предусматривают извлечение факта с использованием нескольких этапов, чтобы избежать выполнения сложного анализа всех представляющих интерес документов. Фактографические описания документов 5 распознаются согласно таблице слов факта на предварительном этапе. Эти фактографические описания могут быть обеспечены признаком их частей речи - существительного или глагола. Далее, на последующем этапе по этим фактографическим описаниям может быть сделан более подробный анализ, чтобы тем 10 самым избежать такого подробного анализа по всем представляющим интерес документам. Для каждого фактографического описания можно определять лингвистические элементы, и затем можно использовать исключения и оценки для удаления фактографических описаний, которые, скорее всего, не являются фактами. Фактографические описания, оставшиеся после исключений и оценки, могут далее 15 быть представлены как факт.

На фиг.1 изображен пример компьютерной системы 100, которая обеспечивает операционную среду для вариантов осуществления. Изображенная компьютерная система 100 может быть стандартной, универсальной программируемой 20 компьютерной системой 100, включающей в себя процессор 102, а также различные компоненты, в том числе массовую память 112, память 104, адаптер 108 дисплея и одно или несколько устройств 110 ввода, например клавиатуру, вспомогательную клавиатуру, мышь и т.п. Процессор 102 обменивается информацией с каждым из компонентов через шину 106 передачи данных. Компьютерная система 100 также 25 может включать в себя сетевой интерфейс 124, например проводное или беспроводное соединение, которое обеспечивает компьютерной системе 100 возможность обмениваться информацией с другими компьютерными системами через сети передачи данных. Компьютерная система 100 может в качестве альтернативы быть жестко 30 закодированным специализированным устройством, которое реализует один или несколько вариантов осуществления.

В примере на фиг.1 процессор 102 реализует команды, сохраненные в массовой памяти 112 в виде операционной системы 114. Операционная система 114 в этом примере обеспечивает основу, на которой могут быть реализованы различные 35 приложения, использующие компоненты компьютерной системы 100. Компьютерная система 100 может реализовать поисковый механизм 118 или аналогичное приложение для обнаружения электронных документов, относящихся к конкретной ситуации. Например, поисковый механизм 118 может принимать термы поиска, введенные 40 непосредственно через устройство 110 ввода пользователем компьютерной системы 100, или может принимать термы поиска, указанные пользователем удаленного компьютера, которые принимаются через сетевой интерфейс 122.

Поиск и/или извлечение факта могут иметь место в отношении одного или нескольких наборов электронных документов, которые содержат текстовую 45 информацию, например, web-страницы, стандартные документы обработки текстов, электронные таблицы и т.д. Эти электронные документы могут быть сохранены локально как набор 116 электронных документов. Эти электронные документы также могут быть сохранены не локально, например, на сетевом запоминающем 50 устройстве 124, содержащем набор 126 электронных документов. Сетевое запоминающее устройство 124 представляет запоминающее устройство локальной сети, контролируемые центральным процессором ячейки памяти в Интернете и т.д. Сетевое запоминающее устройство 124 доступно через сетевой интерфейс 122.

Кроме того, эти варианты осуществления обеспечивают логику для реализации процессором 102 для извлечения фактов из электронных документов 116, 126.

Инструмент 120 для извлечения факта может находиться на локальном запоминающем устройстве 112 как компонент операционной системы 114, или как компонент поискового механизма 118, или как другое приложение, или как автономное приложение, которое может формировать свои собственные независимые результаты. Логические операции, выполняемые вариантами осуществления инструмента 120 для извлечения факта, обсуждаются ниже согласно фиг.2 - фиг.5.

Компьютерная система 100 по фиг.1 может включать в себя множество машиночитаемых носителей информации. Такие машиночитаемые носители информации содержат команды для работы компьютерной системы и для реализации вариантов осуществления, обсуждаемых в этом документе. Машиночитаемыми носителями информации могут быть любые доступные носители информации, к которым компьютер 100 может получить доступ и которые содержат как энергозависимые, так и энергонезависимые носители информации, съемные и несъемные носители информации. В качестве примера, машиночитаемые носители информации могут включать в себя компьютерные носители информации, средства связи и т.д.

Компьютерные носители информации включают в себя энергозависимые и энергонезависимые, съемные и несъемные носители информации, реализованные любым способом или технологией для хранения информации, например, машиночитаемые команды, структуры данных, программные модули или другие данные. Компьютерные носители информации включают в себя, например, RAM, ROM, EEPROM, флэш-память или другую технологию памяти, CD-ROM, универсальные цифровые диски (DVD) или другой накопитель на оптических дисках, магнитофонные кассеты, магнитную ленту, накопитель на магнитных дисках или другие магнитные запоминающие устройства, или любой другой носитель информации, который можно использовать для хранения требуемой информации и к которому компьютерная система 100 может получить доступ.

Коммуникационная среда обычно воплощает машиночитаемые команды, структуры данных, программные модули или другие данные в модулированном сигнале данных, таком как несущее колебание или другой транспортный механизм, и включают в себя любые среды доставки информации. Термин "модулированный сигнал данных" означает сигнал, одна или более характеристик которого установлены или изменяются таким образом, чтобы кодировать в этом сигнале информацию. В качестве неограничивающего примера, коммуникационная среда включает в себя проводную среду передачи, такую как проводная сеть или прямое проводное соединение, и беспроводную среду передачи, такую как акустическая, радиочастотная, инфракрасная и другие беспроводные среды. Комбинации любых приведенных выше носителей и сред также должны включаться в объем машиночитаемых носителей информации.

На фиг.2 изображен пример логических операций, выполняемых поисковым механизмом 118 вместе с инструментом 120 для извлечения факта. В этом примере инструмент 120 для извлечения факта используется до поиска, осуществляемого для формирования библиотеки фактов, присутствующих в электронных документах, в которых будет осуществляться поиск. Следовательно, не требуется время обработки для извлечения фактов, вместо этого эти факты уже извлечены и выбраны из библиотеки фактов на основе введенных термов поиска.

Логические операции начинаются с операции 202 сбора, в ходе которой получают набор электронных документов или каким-либо другим способом осуществляют доступ. Например, электронные документы, в которых со временем будет осуществляться поиск, могут быть сохранены на локальном запоминающем устройстве или могут запрашиваться для доступа по сети. Далее с каждым из этих электронных документов работает инструмент 120 для извлечения факта, который пытается извлечь все факты, которые присутствуют в упомянутых электронных документах. Инструмент 120 для извлечения факта может формировать библиотеку фактов, которые сохраняют совместно с соответствующими электронными документами и которые доступны во время будущих поисков. Например, такая библиотека ассоциаций представлена в таблице 1.

Таблица 1	
Электронный документ	Факты
www.sample1.com	Факт А Факт В Факт С
www.sample2.com	Факт АА Факт ВВ Факт СС
www.sample3.com	Факт ААА

Продолжая последовательность операций фиг.2, пользователь, которому требуется выполнить поиск для обнаружения соответствующих электронных документов, и, в частности, найти соответствующие факты из этих электронных документов, вводит терм поиска в поисковый механизм 118 при операции 206 с термом. В этом примере поисковый механизм 118 далее осуществляет поиск в упомянутых электронных документах на предмет термов поиска и находит соответствующие документы при операции 208 с документами. Поисковый механизм также находит ранее извлеченные факты, которые соответствуют термам поиска, из этих соответствующих электронных документов, и затем выводит на экран соответствующие документы или ссылку на них вместе с соответствующими фактами при операции 210 вывода на экран. Например, терм поиска может быть найдено в www.sample1.com, и также может быть обнаружено, что этот терм поиска соответствует Факту А и Факту В так, что на экран выводится ссылка на www.sample1.com вместе с Фактом А и Фактом В. Соответственно, пользователю быстро предоставляются факты, относящиеся к термам поиска, которые были введены. Пример такого экранного устройства отображения описан ниже со ссылкой на фиг.6.

Конечно, в качестве альтернативы поиск может осуществляться только в отношении ранее извлеченных фактов, а не в отношении самих электронных документов. Кроме того, при определенных обстоятельствах ранее извлеченные факты могут соответствовать термам поиска независимо от того, соответствуют ли электронные документы, содержащие упомянутые факты, термам поиска.

На фиг.3 изображен другой пример логических операций, выполняемых поисковым механизмом 118 вместе с инструментом 120 для извлечения факта. В этом примере инструмент 120 для извлечения факта используется во время поиска для обнаружения фактов, присутствующих в электронных документах, по мере их обнаружения при поиске. Следовательно, нет необходимости в извлечении фактов предварительного поиска и нет необходимости в хранении библиотеки фактов. При таком сценарии инструмент для извлечения факта может бегло просматривать только фрагменты или краткие описания документа для обеспечения очень быстрых результатов, или также

может быть бегло просмотрен весь документ для извлечения всех потенциальных фактов.

Логические операции начинаются с операции 302 с термом поиска, где пользователь вводит терм поиска в поисковый механизм 118. В этом примере поисковый механизм 118 далее осуществляет поиск в упомянутых электронных документах на предмет термов поиска и находит соответствующие документы при операции 304 с документами. После этого инструмент 120 для извлечения используют при операции 306 извлечения для анализа электронных документов, которые были найдены при поиске, для извлечения фактов из тех документов, которые относятся к термам поиска. В результате операции 306 извлечения может формироваться временный набор ассоциаций между электронными документами и фактами, как представлено в таблице 1, который может затем быть помещен в постоянное запоминающее устройство в ожидании последующих поисков по этим термам поиска. После этого при операции 308 вывода на экран поисковый механизм выводит на экран соответствующие документы или ссылку на них вместе с соответствующими фактами, возвращенными инструментом 120 для извлечения факта при операции 306 извлечения.

На фиг.4 изображен многоэтапный подход, используемый вариантами осуществления инструмента 120 для извлечения факта. Вначале, инструмент 120 для извлечения факта пытается распознать набор фактографических описаний из представляющих интерес электронных документов при операции 402 распознавания. Цель данной работы состоит в том, чтобы найти те описания в тексте, которые, скорее всего, являются фактами, на основе обнаружения соответствий таблице слов факта, более подробно обсуждаемой ниже со ссылкой на фиг.5. С выполнением быстрой установки соответствия большая часть электронного документа, которая должна быть проигнорирована при поиске фактов, может быть удалена из дальнейшей обработки для извлечения факта, тем самым увеличивается эффективность последующего(их) этапа(ов), которые используются для увеличения точности.

После идентификации набора фактографических описаний для анализируемого документа далее при операции 404 извлечения в этом наборе фактографических описаний выполняется извлечение факта. Здесь более подробный анализ выполняется только в наборе фактографических описаний, в отличие от целого документа, для поддержки удовлетворительной эффективности при достижении адекватной точности. Анализ операции извлечения включает в себя принятие решения на основе определения лингвистических элементов фактографических описаний. Такие лингвистические элементы могут включать в себя синтаксические элементы, семантику и т.д.

На фиг.5 изображен пример подробностей операций распознавания и извлечения фиг.4. Логические операции начинаются с операции 502 сканирования, где инструмент 120 для извлечения факта сканирует электронный документ для обнаружения слов или словосочетаний, соответствующих словам или словосочетаниям из таблицы слов факта. Таблица слов факта является списком слов или словосочетаний, которые, как известно, скорее всего используются при выражении факта, в отличие, например, от мнения. В таблице 2 представлен короткий пример. Следует отметить, что для обеспечения выполнения оптимальной обработки слова из этой таблицы могут быть обеспечены признаком наиболее подходящей части речи (POS), который описан ниже согласно операции 504 с признаком.

Список слов факта	Признаки POS
Слово/Словосочетание 1	Признак POS
Слово/Словосочетание 2	Признак POS
Слово/Словосочетание N	Признак POS

5

Для определения слов, при которых приходит мысль о фактах, а не о мнениях, было проведено исследование. Например, класс слов, которые представляют факты, может быть получен с использованием исследования и работы по классификации глаголов и их лексических функций. Две соответствующих работы, которые можно использовать как материал для этого, включают в себя:

10

(1) Mel'cuk (1996) *Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon*. In L. Wanner (ed.): *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia: Benjamins, 37-102.

15

(2) Fontenelle, T. (1997): "Discovering Significant Lexical Functions in Dictionary Entries", in Cowie, AP. (ed.) *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, Oxford.

20

Соответственно, на основе такого исследования может быть создан список слов факта, как показано в таблице 2, включающий в себя эти глаголы или другие слова, которые наводят на мысль о выражении факта, в отличие от нефактической информации. Например, термины поиска "изобретенный" или "нанятый" наводят на мысль о выражении факта, тогда как термины "может быть" или "жалуется" не наводят на подобную мысль. Конкретный пример списка слов факта может быть найден в приложении А, находящемся в конце этого описания. Этот конкретный пример не является исчерпывающим списком глаголов, которые являются словами факта и могут использоваться для обнаружения фактографических описаний в электронных документах.

25

30

После применения таблицы слов факта к электронному документу или параллельно с применением таблицы слов факта, например, где признак POS уже связан со словами в таблице слов факта, части речи (POS) каждого из слов каждого фактографического описания обеспечиваются признаком при операции 504 обеспечения признаком. Эта операция 504 обеспечения признаком, которая может выполняться параллельно с операцией 502 сканирования (беглого просмотра) или после нее, может включать в себя устранение неоднозначности выборов для слов, которые имеют несколько признаков POS, например, предпочтение признака существительного признаку глагола, так как подразумевается, что синтаксические словосочетания, подобные именным словосочетаниям, как известно, являются сущностями, связанными с фактографическим событием. По этой причине любые неизвестные и предварительно не обеспеченные признаками слова могут также по умолчанию объявляться существительными. Как и существительные, прилагательные могут быть предпочтительнее глаголов (например, "запланированный" как прилагательное по сравнению с "запланирован" как глагол), а также те слова, которые имеют признак и прилагательного и глагола, по умолчанию будут объявляться прилагательными, так как прилагательное является частью именных словосочетаний, которые, как известно, являются сущностями, связанными с фактографическим событием. При создании ассоциаций признаков POS со словами таблицы слов факта, например, при создании таблицы, эти выборы с устранением неоднозначности могут уже быть применены, так что, например, слово "запланированный" в таблице связано с признаком POS прилагательного, а не с признаком POS глагола.

35

40

45

50

После обнаружения фактографических описаний и обеспечения слов

фактографических описаний признаками POS может быть выполнен более полный анализ для улучшения точности извлечения факта без требования применения этой более полной обработки ко всему документу. При операции 506 идентификации идентифицируются синтаксические словосочетания, подобные именным словосочетаниям и глагольным словосочетаниям. Синтаксические словосочетания идентифицируются с использованием общепринятых правил грамматики и несложного лингвистического анализа. Идентифицируются окружающие синтаксические словосочетания, т.е. находящиеся в самой непосредственной близости от набора фактографических описаний в документе, и если у фактографического описания нет связанных с ним синтаксических словосочетаний, то соответствующее предложение может быть удалено из дальнейшего рассмотрения. Соответственно, при сосредоточении только на тех синтаксических словосочетаниях, которые находятся в окружении фактографического описания, избегают просмотра всех лингвистических элементов целого предложения.

Кроме того, при операции 506 идентификации с оценкой роли, которую синтаксическое словосочетание играет в соответствующем предложении, на основе образца, идентифицированного в фактографическом описании, далее определяются лингвистические элементы фактографических описаний, имеющих окружающие синтаксические словосочетания. Соответственно, из образца слова фактографического описания определяется, играет ли синтаксическое словосочетание роль подлежащего или дополнения в предложении, содержащем анализируемое в данный момент фактографическое описание.

После определения лингвистических элементов фактографических описаний, т.е. когда синтаксические словосочетания и их роли идентифицированы, далее при операции 508 исключения к этим именным словосочетаниям фактографических описаний могут быть применены правила исключения для дальнейшего удаления тех, которые, скорее всего, не являются выражением факта. Правила исключения могут применяться на основании того, что синтаксическое словосочетание является дополнением, синтаксическое словосочетание является подлежащим или без учета роли синтаксического словосочетания. Кроме того, в этом конкретном варианте осуществления правило исключения, применяемое к отдельным словам, к синтаксическим словосочетаниям или к целому предложению, приводит к идентичному результату, который заключается в исключении целого предложения из фактографических описаний. Пример правил исключения, которые могут быть применены, представлен в таблице 3.

Таблица 3	
Правила исключения	Вывод
«Дополнение» имеет модификатор «мнение/необъективный»	Исключить предложение-кандидат
Фильтры предложения: - начальное слово предложения (например, местоимения) - пунктуация: например «?»	Исключить предложение-кандидат
«Подлежащее» имеет определенный артикль - если не имя собственное	Исключить предложение-кандидат
Окружающий «Контекст» «Дополнения»	Исключить предложение-кандидат, если окружающий контекст имеет конкретный POS, который не указывает на факт (например, некоторый класс местоимений)
В предложении встречаются стоп-слова	Исключить предложение-кандидат
«Подлежащее» (или) «Объект» содержат местоимения	Исключить именное словосочетание

После применения правил исключения или параллельно с применением правил исключения применяют правила оценки при операции 510 оценки. Правила оценки

задают вес и именным словосочетаниям-подлежащим, и именным словосочетаниям-дополнениям для каждого из различных признаков, и общей оценкой для фактографического описания-кандидата является сумма весов отдельных признаков плюс оценка достоверности соответствующего слова факта. Веса отдельных признаков могут быть положительными при указании на факт и могут быть отрицательными при указании на нефактическую информацию. Примеры признаков и связанных правил оценки приведены ниже в таблице 4. Оценки признаков могут назначаться вручную с использованием суждения человека, или их можно узнавать автоматически.

Таблица 4	
Признаки	Правила оценки
Оценка достоверности соответствующего образца (слово факта, например, смысловой глагол)	
Класс ролей (т.е. подлежащее или глагол), например человек, страна, организация и т.д.	Оценка для каждого класса
Основное «Подлежащее» содержит имя собственное	Обычный вес
Длина «Дополнения»	Оценка длины
Длина «Подлежащего»	Оценка длины
Длина предложения	Оценка длины
«Подлежащее» появляется в начале предложения - например, вынос подлежащего «Дополнение» имеет модификатор (прилагательное, наречия)	Положительная оценка Отрицательный - Основной вес
«Дополнение» имеет определенный артикль ("the")	Отрицательный - Основной Исключается, когда заканчивает предложение с глаголом-связкой

Далее, при операции 512 вопроса общая оценка для фактографического описания сравнивается с предопределенным порогом для определения, превышает ли общая оценка порог. Если порог не превышен, то соответствующее фактографическое описание может быть отвергнуто. Если порог превышен, то фактографическое описание, полное предложение и/или полный абзац или другая часть документа могут быть представлены как факт при операции 514 представления. Это представление может включать в себя вывод на экран факта, сохранение факта в библиотеке и т.д.

При использовании правил оценки и сравнении с порогом, весами, назначенными признакам, и/или значением порога можно манипулировать без манипуляции целым подходом к извлечению факта. Следовательно, степенью точности извлечения факта и представлением можно управлять, в то время как этапы обработки остаются неизменными.

На фиг.6 изображен иллюстративный снимок 600 экрана, получающийся в результате выполнения поиска. Термы поиска были введены в поисковое поле 602 для проведения поиска. Терм поиска был сопоставлен с различными ссылками 604 на web-сайты, доступные в Интернете. Пользователь может обращаться к электронным документам обычным способом.

Кроме того, факты 610, 612 и 614 о терме поиска выводятся на экран в разделе 608. Соответственно, пользователь может быстро определять факты о предмете поиска без необходимости обращения к какому-либо из электронных документов, которые были найдены, и без необходимости самому читать и отличать факт от мнения. В этом конкретном примере факты 610, 612 и 614 включают в себя гиперссылки, которые пользователь может выбирать, для предоставления подробной информации об источнике факта и/или для представления контекста, в котором факты были обнаружены (например, связанная с фактом дата, другие факты и т.д.).

Следует понимать, что снимок 600 экрана является лишь одним примером того, как факты могут быть представлены пользователю. Вместо того, чтобы представлять их в отдельном столбце, как показано, они могут быть перечислены как подэлементы электронного документа, из которого они были извлечены. Кроме того, в качестве альтернативы перечислению фактов на странице с результатами поиска или в дополнение к нему, факты, извлеченные из конкретного электронного документа, могут также быть перечислены в столбце или другом месте при просмотре пользователем самого электронного документа. Кроме того, в качестве альтернативы отделению фактов от документа для вывода на экран или в дополнение к нему, факты могут выделяться в электронных документах и в списке 604 документов в пределах результатов поиска, и в пределах всего электронного документа, когда его выбирают для вывода на экран. В качестве еще одной альтернативы, факты могут выводиться на экран независимо от результатов поиска, например, вывод на экран только фактов с выбираемой ссылкой для получения исходных документов, где осуществлен поиск только извлеченных фактов, чтобы тем самым полностью избежать поиска в документе.

Кроме того, следует понимать, что представление извлеченных фактов, например представленных на снимке 600 экрана, может быть обеспечено в виде вывода на экран локального компьютера с реализацией поиска и извлечения факта для локального пользователя. В качестве альтернативы, представление извлеченных фактов, например изображенных на снимке 600 экрана, может быть обеспечено в виде вывода на экран удаленного компьютера, который запросил локальный компьютер выполнить поиск и извлечение факта от его имени, например, в случае поискового механизма, размещенного в Интернете.

Соответственно, можно эффективно и точно извлекать факты из документов для представления пользователям. За счет многоэтапного подхода увеличена эффективность при исключении необходимости детализированного анализа всех документов, а также при исключении необходимости детализированного анализа всего предложения, где было найдено фактографическое описание. Точность поддерживается с использованием дополнительного анализа фактографических описаний, которые были обнаружены в документе на предварительном этапе обработки.

Несмотря на то что изобретение было изображено и описано, в частности, согласно различным вариантам его осуществления, специалистам в данной области техники будет понятно, что могут быть внесены различные другие изменения по форме и в деталях, не выходящие за пределы сущности и объема изобретения. Например, при проведении синтаксического анализа фактографического описания, а не позже, во время применения других правил исключения, могут быть применены определенные правила исключения, которые не имеют отношения к лингвистическим элементам фактографического описания, например правила исключения, основанные на пунктуации предложения.

Приложение А - Слова факта

abase (унижать) abate (уменьшать) abort (прерывать) abrade (стирать) abridge (сокращать) absorb (поглощать) abstract (абстрагировать) accelerate (ускорять) accent (акцентировать) accept (принимать) accredit (аккредитовать) achieve (достигать) act (действовать) add (добавлять) address (обращаться) adduce (представлять) adjust (регулировать) administer (управлять) admit (признавать) advance (продвигать) advertise (рекламировать) aerate (проветривать) afford (предоставлять) aggravate (ухудшать)

agree (соглашаться) aid (помогать) aim (нацеливать) air (проветривать) allay (смягчать)
 alleviate (облегчать) alter (изменять) amend (исправлять) amplify (усиливать) amuse
 (развлекать) animate (оживлять) announce (объявлять) answer (отвечать) antedate
 (предшествовать) appear (появляться) appease (успокаивать) apply (применять) argue
 5 (утверждать) arouse (пробуждать) arrange (организовывать) arrest (арестовывать) arrive
 (прибывать) ask (спрашивать) assemble (собирать) assert (утверждать) asseverate
 (торжественно заявлять) assign (назначать) assuage (успокаивать) assure (уверять)
 attach (прилагать) attack (нападать) attenuate (уменьшать) avert (предотвращать) avoid
 10 (избегать) awake (будить) award (награждать) back (поддерживать) bail (брать на
 поруки) bank (наваливать) bar (преграждать) barbarize (одичать) bare (обнажать) base
 (базировать) batter (разбивать) beach (вытаскивать на берег) beam (излучать) bear
 (переносить) become (становиться) befog (затуманивать) befuddle (удивлять) beget
 (порождать) begin (начинать) begrime (чернить) belch (изрыгать) belie (противоречить)
 15 bend (сгибать) benumb (парализовать) bequeath (завещать) bestow (даровать) betray
 (предавать) better (улучшать) bind (связать) blackleg (жульничать) blanket (покрывать)
 bleach (отбеливать) blemish (портить) blend (смешивать) blight (разрушать) blister
 (вызывать пузыри) block (блокировать) blockade (блокировать) blow (дуть) blunder
 20 (натыкаться) blunt (притуплять) blur (размывать) blurt (выбалтывать) bob (слегка
 ударять) bog (увязать) boil (кипятить) bolster (поддерживать) boost (повышать)
 bowdlerize (выхолащивать) bowl (катать) brace (окружать) brand (клеймить) brave
 (выдерживать) break (ломать) brief (сокращать) brighten (проясняться) bring
 (приносить) broadcast (вещать) bruise (ушибать) buckle (скреплять пряжкой) build
 25 (строить) bull (играть на повышение) bunch (связывать) bundle (связывать) bung
 (закупоривать) burlesque (пародировать) burn (жечь) burst (разрывать) bury (хоронить)
 buy (покупать) bypass (обходить) canvass (агитировать) cap (увенчивать) capitalize
 (капитализировать) carry (нести) cast (бросать) castigate (наказывать) castrate
 30 (кастрировать) catch (ловить) chafe (тереть) change (изменять) channel (направлять)
 charge (обвинять) check (проверять) chill (охлаждать) chime (звенеть) chip (разбиваться)
 chock (подпирать) choke (душить) choose (выбирать) churn (взбалтывать) cipher
 (шифровать) circulate (циркулировать) circumvent (обходить) claim (требовать) clash
 (сталкиваться) clean (чистить) cleanse (чистить) clear (очищать) climb (подниматься)
 35 clinch (договариваться) clip (обрезать) clog (засоряться) close (закрывать) clot
 (сгущаться) cloud (омрачать) cogle (морщиться) coin (чеканить) collapse (разрушаться)
 collect (собирать) colour (окрашивать) comfort (успокаивать) commission
 (уполномочивать) commit (совершать) communicate (сообщать) compare (сравнивать)
 40 complete (заканчивать) compound (составлять) compress (сжимать) compromise
 (компрометировать) conceal (скрывать) concede (уступать) conceive (понимать)
 conciliate (примирять) conclude (заключать) conduct (проводить) confess (признаваться)
 confide (доверять) confirm (подтверждать) confound (путать) confuse (путать) congeal
 (замораживать) connect (соединять) conserve (сохранять) consolidate (объединять)
 45 constitute (составлять) constrain (ограничивать) constrict (сжимать) continue
 (продолжать) contort (искажать) contract (сокращать) control (управлять) convert
 (преобразовать) convey (передавать) cook (готовить) cool (охлаждать) cordon
 (загораживать) correct (исправлять) corrode (разъедать) corrupt (развращать) counter
 50 (противостоять) countersink (зенковать) cover (покрывать) crack (взламывать) crank
 (проворачивать) crash (разбивать) craze (сводить с ума) create (создавать) cripple
 (калечить) crop (подрезать) cross (пересекать) crumble (крошить) crush (давить) cry
 (кричать) curb (обуздывать) curdle (свертываться) curtail (сокращать) cushion

(смягчать) cut (резать) damage (повреждать) damp (заглушать) dance (танцевать) dangle
 (свисать) darken (темнеть) darn (чинить) dash (разбивать) deaden (ослаблять) deal (иметь
 дело) debase (понижать качество) debauch (соблазнять) debunk (разоблачать) decay
 (разлагать) decide (решать) declare (объявлять) deepen (углублять) deface (стирать)
 5 defeat (побеждать) defend (защищать) deflate (выкачивать) deflect (отклонять) deform
 (искажать) defrost (размораживать) delay (задерживать) delegate (делегировать) deliver
 (поставлять) demise (передавать по наследству) demonstrate (демонстрировать) dent
 (вдавливает) deny (отрицать) deplete (исчерпывать) depreciate (обесценивать) depress
 10 (подавлять) deprive (лишать) depute (передавать) derange (нарушать) describe
 (описывать) desecrate (осквернять) design (проектировать) designate (обозначать)
 desolate (опустошать) despoil (грабить) destroy (разрушать) detail (детализировать)
 detect (обнаруживать) deteriorate (ухудшать) determine (устанавливать) develop
 (развивать) die (умирать) differentiate (дифференцировать) diffuse (разбрасывать) dilute
 15 (растворять) dim (тускнеть) diminish (уменьшать) direct (направлять) dirty (пачкать)
 disable (калечить) disappear (исчезать) discharge (разгружать) discipline
 (дисциплинировать) disclose (раскрывать) discolour (обесцвечивать) disconnect
 (разъединять) discontinue (прекращать) discover (обнаруживать) discuss (обсуждать)
 20 disfigure (уродовать) disguise (маскировать) dislocate (сместить) dislodge (сместить)
 dismantle (демонтировать) dismount (снимать) disorder (приводить в беспорядок)
 dispatch (отправлять) dispense (распределять) disperse (рассеивать) display (отображать)
 dispute (дискутировать) disrupt (разрушать) distil (дистиллировать) distinguish
 (различать) distort (искажать) disturb (нарушать) divert (отклонять) divide (делить) dock
 25 (состыковывать) doctor (лечить) dodge (избегать) double (удваивать) douse (окунает)
 draft (проектировать) dramatize (драматизировать) draw (тянуть) dredge (посыпать)
 dress (украшать) drive (ездить) drop (понижать) drown (тонуть) duff (подновлять) dull
 (притуплять) earth (закапывать) ease (ослаблять) eat (есть) educate (обучать) effect
 30 (осуществлять) elevate (поднимать) elicit (выявлять) elude (уклоняться) emancipate
 (эмансипировать) embellish (украшать) embitter (озлоблять) embody (воплощать) emit
 (испускать) emphasize (подчеркивать) enable (позволять) encourage (поощрять) end
 (заканчивать) endorse (подтверждать) endow (обеспечивать) enforce (заставлять) engage
 (участвовать) enhance (увеличивать) enjoin (предписывать) enlarge (увеличивать)
 35 enliven (оживлять) ennoble (облагораживать) enrich (обогащать) enrol (регистрировать)
 enshrine (хранить) entail (влечет за собой) entangle (запутывать) enthrone (возводит на
 престол) entrust (поручать) enunciate (излагать) epitomize (воплощать) equalize
 (уравнивать) erect (устанавливать) escalate (наращивать) establish (устанавливать) evade
 40 (уклоняться) evaporate (испаряться) evince (проявлять) evoke (вызывать) exacerbate
 (усиливать) exact (взыскивать) exaggerate (преувеличивать) examine (исследовать)
 exasperate (сердить) exceed (превышать) excite (возбуждать) exhale (выдыхать) exhibit
 (выставлять) exist (существовать) expand (расширять) expedite (ускорять) explain
 (объяснять) expose (выставлять) expound (разъяснять) express (выразить) extend
 45 (расширять) extinguish (гасить) extort (вымогать) extract (выделять) fabricate
 (изготавливать) face (встречаться) fade (исчезать) fail (подводить) fake (подделывать) fall
 (падать) falsify (фальсифицировать) familiarize (ознакомлять) fasten (закреплять) father
 (порождать) fatten (откармливать) feature (характеризовать) feed (кормить) ferry
 50 (переправлять) fertilize (оплодотворять) festoon (украшать) fiddle (играть) fight
 (бороться) fill (заполнять) filter (фильтровать) finalize (завершать) find (находить)
 finish (заканчивать) fire (стрелять) fit (приспосабливать) fix (закреплять) flag
 (сигнализировать) flash (высвечивать) flaunt (щеголять) flay (свежевать) float (плавать)

flood (затоплять) floodlight (освещать прожектором) flourish (процветать) flush
 (смыть) fly (летать) fog (затуманивать) foil (мешать) fold (сгибать) follow (следовать)
 force (вынуждать) forge (ковать) forgive (прощать) form (формировать) foster
 (способствовать) foul (загрязнять) found (основанный) frame (обрамлять) fray
 5 (изнашиваться) free (освобождать) freeze (замораживать) frustrate (разбивать) furl
 (сворачивать) furnish (снабжать) furrow (пахать) fuse (плавить) gain (приобретать)
 gallop (скакать) garble (искажать) gash (наносить рану) generate (производить)
 gerrymander (манипулировать фактами) get (получать) give (давать) gladden (радовать)
 10 glorify (прославлять) gloss (придавать блеск) glut (насыщать) go (идти) govern
 (управлять) grade (сортировать) graduate (градуировать) grant (предоставлять) grate
 (тереть) graze (задевать) ground (обосновывать) group (группировать) grow (расти)
 guide (руководствовать) halt (останавливать) halve (сокращаться наполовину) hamper
 15 (препятствовать) handle (обрабатывать) happen (случаться) harass (беспокоить) harbour
 (стать на якорь) harden (укреплять) harm (вредить) harmonize (гармонизировать) harry
 (изматывать) hasten (спешить) hatch (штриковать) head (возглавлять) heal (излечивать)
 hear (слышать) heat (нагревать) heighten (усиливать) help (помогать) hide (скрывать) hit
 (ударять) hoard (копить) hoist (поднимать) hold (держать) hope (надеяться) hound
 20 (преследовать) hurt (повреждать) identify (идентифицировать) illuminate (освещать)
 imagine (воображать) impair (ослаблять) impart (обеспечивать) impeach (привлекать к
 ответственности) impede (препятствовать) imperil (подвергать опасности) implant
 (внедрять) improve (улучшать) inaugurate (вводить в должность) increase (увеличивать)
 25 indent (зазубривать) indenture (связывать договором) indicate (указывать) induce
 (побуждать) induct (вводить в должность) infect (заражать) infiltrate (пропитывать)
 infix (вставляя) inflame (воспламенять) inflate (раздувать) inflict (причинять) influence
 (влиять) inform (сообщать) infuse (вселять) initial (подписывать) initiate
 (инициализировать) injure (ранить) insert (вставлять) inspire (вдохновлять) instigate
 30 (провоцировать) instil (прививать) institute (назначать) integrate (объединять) intend
 (предназначать) intensify (усиливать) interpolate (интерполировать) interrupt
 (прерывать) intimate (сообщать) introduce (представлять) invert (инвертировать)
 invigorate (поддерживать) invite (приглашать) invoke (вызывать) involve (вовлекать)
 issue (выпускать) jab (тыкать) jam (зажимать) jettison (выбрасывать за борт) jingle
 35 (звенеть) join (соединять) jumble (смешивать) jump (подскакивать) justify
 (оправдывать) keep (продолжать) kick (пинать) kill (убивать) kindle (разжигать) knock
 (стучать) lacerate (раздирать) ladder (взбираться по лестнице) lance (метать) land
 (приземляться) laugh (смеяться) launch (запускать) lay (класть) layer (наслаивать) lead
 40 (приводить) leave (оставлять) lend (предоставлять) lengthen (удлинять) lessen
 (уменьшать) let (позволять) level (выравнивать) liberate (освобождать) lie (лежать) light
 (освещать) lighten (освещать) limit (ограничивать) line (выравнивать) link (связывать)
 listen (слушать) litter (сорить) live (жить) liven (оживлять) load (загружать) lock
 (запирать) loose (освобождать) loosen (ослаблять) lose (терять) lower (понижать) lump
 45 (образовывать комки) magnify (увеличивать) maintain (поддерживать) make (делать)
 manage (управлять) mangle (корезить) manipulate (манипулировать) manufacture
 (производить) mark (отмечать) marshal (выстраивать) mask (маскировать) match
 (соответствовать) matter (иметь значение) maul (расщеплять) measure (измерять) meet
 50 (встречать) mellow (созревать) melt (таять) mend (исправлять) mention (упоминать)
 mildew (поражать милдью) mind (возражать) misrepresent (искажать) miss (пропускать)
 mist (затуманиваться) mitigate (смягчать) modify (изменять) mollify (успокаивать) moot
 (обсуждать) mould (формировать) move (перемещать) muddle (запутывать) muddy

(пачкать) muffle (приглушать) muss (приводить в беспорядок) muster (собирать) mute
 (приглушать) mutilate (искажать) narrow (сужать) navigate (осуществлять навигацию)
 neaten (убирать) nick (отмечать) nip (прищемить) notch (маркировать) notice (обращать
 5 внимание) nourish (кормить) nurse (нянчить) obfuscate (запутывать) obscure (затенять)
 obstruct (затруднять) obtain (получать) occur (занимать) occur (происходить) offend
 (оскорблять) offer (предлагать) open (открывать) operate (управлять) oppose
 (возражать) order (заказывать) originate (происходить) outline (обрисовывать в общих
 чертах) overcharge (перегружать) overdo (переусердствовать) overflow (переполнять)
 10 overturn (опрокидывать) overwork (переутомлять) pacify (умиротворять) pack
 (упаковывать) pad (дополнять) panic (паниковать) paralyze (парализовать) pare
 (чистить) parlay (поставить на кон) parole (условно освободить) parry (парировать)
 part (разделять) partition (разделить) pass (передавать) patch (исправлять) pay (платить)
 real (трезвонить) peddle (торговать вразнос) peg (привязывать) penalize (штрафовать)
 15 perform (выполнять) perish (погибать) persecute (преследовать) pervert (извращать)
 phrase (формулировать) pick (выбирать) pillow (служить подушкой) pique (задевать) pit
 (делать ямки) placard (использовать плакаты для рекламы) place (размещать) plan
 (планировать) plant (сажать) play (играть) pluck (щипать) plug (включать) plunge
 20 (погружать) point (указывать) poison (отравлять) pole (подпирать шестами) polish
 (полировать) poll (опрашивать) pool (объединять) pop (хлопать) pose (позировать)
 position (помещать) post (отправлять) pound (загонять) preach (проповедовать)
 precipitate (ускорять) predate (предшествовать) prefer (предпочитать) prejudice (наносить
 ущерб) preoccure (занимать) prepare (готовить) present (представлять) preserve
 25 (сохранять) prettify (украшать) prevent (предотвращать) prick (укалывать) prime
 (заправлять) proclaim (объявлять) procure (обеспечивать) produce (производить) profess
 (выражать) programme (программировать) promote (продвигать) promulgate
 (провозглашать) prop (подпирать) propagandize (пропагандировать) propel
 30 (продвигать) propound (представлять на обсуждение) prosecute (преследовать по суду)
 protect (защищать) protest (возражать) prove (доказывать) provide (обеспечивать)
 provoke (вызывать) prune (сокращать) publicize (разглашать) publish (издавать) pull
 (тянуть) pulp (превращать в мягкую массу) punch (избивать) puncture (прокалывать)
 punish (наказывать) punt (плыть на плоскодонке) purge (очищать) push (толкать) put
 35 (помещать) qualify (квалифицировать) quarter (квартировать) quench (подавлять)
 question (расспрашивать) quicken (ускорять) quieten (успокаивать) quilt (стегать) race
 (мчаться) raise (поднимать) ransack (рыться) rap (стучать) rationalize
 (рационализировать) rattle (грохотать) re-engage (вновь сцеплять) re-establish
 40 (восстанавливать) re-form (преобразовывать) read (читать) rear (поднимать) reawaken
 (снова пробуждать) recall (вспоминать) receive (принимать) reclaim (востребовать)
 recline (откидывать) recognize (признавать) recommend (рекомендовать) reconcile
 (примирять) reconsider (пересматривать) record (записывать) recruit (вербовать) reduce
 (уменьшать) refer (ссылаться) refine (очищать) reflect (отражать) refloat (заново
 45 выпускать заем) reform (преобразовывать) refuse (отказывать) regard (расценивать)
 register (регистрировать) regulate (регулировать) rehabilitate (реабилитировать) rehearse
 (репетировать) reinforce (укреплять) reissue (переиздавать) reject (отклонять) rekindle
 (разжигать) relate (иметь отношение) relax (расслаблять) release (освободить) relieve
 50 (освободить) reline (заменять обшивку) remould (восстанавливать протектор) remove
 (удалять) rend (отрывать) renew (возобновлять) renovate (ремонтировать) reopen (вновь
 открывать) repair (восстанавливать) replace (заменять) report (сообщать) republish
 (переиздавать) require (требовать) rerun (запускать повторно) reseal

(переустанавливать) resist (сопротивляться) rest (отдыхать) restart (перезапускать)
 restore (восстанавливать) restrain (ограничивать) result (приводить) resurrect
 (возрождать) retail (продавать в розницу) retain (удерживать) recline (удаляться) retract
 (отрекаться) retrench (сокращать) retrieve (восстанавливать) return (возвращать) reveal
 5 (разоблачать) reverse (полностью изменять) revive (возрождать) rewind (перематывать)
 right (исправлять) ring (звонить) rise (вставать) roast (жарить) rock (качать) roll (катить)
 rotate (вращать) rouse (пробуждать) row (гребсти) ruffle (раздражать) ruin (разрушать)
 rumple (приводить в беспорядок) run (бежать) rush (мчаться) rustle (шелестеть) sail
 10 (плавать) salvage (спасать) sap (иссушать) save (спасать) scald (ошпаривать) scorch
 (опалить) score (подсчитывать) scotch (пресечь) scratch (царапать) scream (кричать)
 scuff (протирать) scurper (топить) scuttle (удирать) seal (запечатывать) sear (иссушать)
 seat (усаживать) secure (обеспечивать) see (видеть) sell (продавать) send (отправлять)
 serve (служить) set (устанавливать) settle (улаживать) sever (разъединять) shake
 15 (встряхивать) shame (позорить) sharpen (заострять) shatter (разрушать) sheathe
 (вкладывать в ножны) shed (проливать) shelter (укрывать) shield (ограждать) shift
 (перемещать) shine (сиять) shingle (крыть) shirk (уклоняться) shoot (стрелять) shorten
 (сокращать) shout (кричать) show (показывать) shrink (сокращать) shut (закрывать) sift
 20 (просеивать) sign (подписывать) signal (сигнализировать) signalize (сигнализировать)
 signify (иметь значение) simmer (кипеть) sing (петь) singe (палить) sink (погружать) sit
 (сидеть) site (помещать) situate (располагать) skirt (окаймлять) slacken (замедлять) slake
 (уменьшать) slash (резать) sleep (спать) slice (нарезать) slip (скользить) slow
 (замедлять) smear (порочить) smile (улыбаться) smudge (оставлять пятна) snag
 25 (поймать) snap (хватать) snarl (спутывать) snuff (нюхать) sober (отрезвлять) soften
 (смягчать) soil (пачкать) solace (сочувствовать) solidify (укреплять) soothe
 (успокаивать) sort (сортировать) sound(звучать) sour (закисать) sow (сеять) spare
 (экономить) spark (зажигать) speak (говорить) speck (усеивать) speed (ускорять) spill
 30 (проливать) spin (прясть) splinter (раскалывать) split (раскалывать) splodge (хлюпать)
 spoil (портить) sponsor (спонсировать) sport (заниматься спортом) spot (определять)
 sprout (извергать) sprain (вывихивать) spray (распылять) spread (распространять) spring
 (снабжать пружиной) square (согласовывать) squash (раздавить) squeeze (сжимать)
 stack (загромождать) staff (укомплектовывать) stain (окрашивать) stalemate (ставить в
 35 безвыходное положение) stall (останавливать) stamp (штамповать) stand
 (выдерживать) star (играть главную роль) starch (крахмалить) start (начинать) staunch
 (останавливать) stay (оставаться) steady (стабилизировать) steer (регулировать) stem
 (происходить) step (ступать) stick (прикреплять) stiffen (напрягаться) still
 40 (успокаивать) stir (размешивать) stoke (топить) stop (останавливать) store (запасать)
 straighten (выправлять) strain (напрягать) strand (переплетать) strengthen (усиливать)
 stress (подчеркивать) stretch (протягивать) strike (ударять) strip (раздевать) strum
 (играть) study (изучать) stuff (наполнять) stultify (сводить на нет) stunt (останавливать
 рост) subdue (подчинять) subscribe (подписывать) subvert (ниспровергать) succeed
 45 (преуспевать) suffer (страдать) suggest (предлагать) suit (удовлетворять) summarize
 (суммировать) supplement (добавлять) supply (поставлять) support (поддерживать)
 suppose (предполагать) suppress (подавлять) surface (появляться) surrender (сдавать)
 survive (переживать) suspend (приостанавливать) sustain (выдерживать) sweep
 50 (смахивать) sweeten (подслащивать) swell (раздувать) swing (качать) swish
 (размахивать) taint (заражать) tarnish (запятнать) task (задавать работу) teach (учить)
 tear (рвать) telephone (звонить) temper (умерять) tend (склоняться) thank (благодарить)
 thaw (таять) thin (утончать) thrill (трепетать) throw (бросать) thrust (толкать) thump

(ударять) thwart (мешать) tidy (приводить в порядок) tighten (уплотняться) toll (звонить) tootle (издавать негромкие звуки) topple (свергать) torment (мучить) torture (мучить) total (насчитывать) touch (касаться) toughen (ужесточать) tousle (ерошить) tow (буксировать) train (обучать) trample (растаптывать) transfer (передавать) transplant (трансплантировать) trap (заманивать в ловушку) travel (путешествовать) treat (лечить) trigger (вызывать) trim (урезать) truss (связывать) try (пробовать) tumble (упасть) turn (поворачивать) twang (звучать) twiddle (вертеть) twirl (вращать) twist (крутить) unblock (открывать) unburden (облегчать бремя) unclog (прочищать) undo (уничтожать) unfasten (откреплять) unfix (откреплять) unfold (разворачивать) unhinge (расстраивать) unhitch (отцеплять) unite (объединять) unloose (ослаблять) unravel (распутывать) unsaddle (расседывать) unseat (сбрасывать) unsex (кастрировать) unstop (откупоривать) untangle (распутывать) untwist (раскручивать) uphold (поддерживать) upset (опрокидывать) urge (убеждать) use (использовать) validate (утверждать) vandalize (разрушать) veer (поворачивать) veil (скрывать) ventilate (проветривать) vocalize (напевать) voice (высказывать) vote (голосовать) vulgarize (опошлять) waft (доноситься) waggle (покачивать) wake (будить) walk (идти) wangle (заполучить) warm (нагревать) warn (предупреждать) warp (деформировать) warrant (гарантировать) wash (вымыть) watch (наблюдать) weaken (ослаблять) wean (отнимать от груди) wear (носить) weave (ткать) weep (плакать) weld (сваривать) whet (точить) whirl (кружить) whitewash (ретушировать) widen (расширять) wield (владеть) wiggle (шевелить) wilt (слабеть) win (выигрывать) wind (проветривать) wing (лететь) wipe (вытирать) wire (телеграфировать) wish (желать) withdraw (отзывать) wither (увядать) withhold (отказывать) work (работать) worry (волновать) wreak (давать выход) wreck (разрушать) wrest (вырывать) wring (скручивать) wrinkle (морщить) write (писать) yield (уступать).

Формула изобретения

1. Способ различения фактов и мнений в электронных ресурсах, осуществляемый с помощью компьютера и выполняемый процессором, причем способ содержит прием термина поиска, содержащего существительное, обнаружение релевантных электронных ресурсов, которые соответствуют терму поиска, отображение списка релевантных электронных ресурсов и фрагментов релевантных электронных ресурсов в списке, содержащих слова, соответствующие терму поиска, сканирование релевантного электронного ресурса для обнаружения фактографических описаний из предложений, которые содержат существительное из термина поиска и один или более глаголов, соответствующих словам таблицы слов факта, построенной так, что она включает в себя список глаголов, определенных как указывающие на выражение факта, удаление из обработки для извлечения фактов тех частей релевантного электронного ресурса, которые содержат слова, не соответствующие терму поиска, и слова таблицы слов факта, исследование обнаруженных фактографических описаний для идентификации лингвистических элементов фактографических описаний после удаления частей релевантного электронного ресурса, определение, следует ли представить фактографическое описание как факт, на основе идентифицированных лингвистических элементов, и

представление по меньшей мере части предложения, которое содержит термы поиска и фактографическое описание, определенное как факт, релевантный терму поиска.

2. Способ по п.1, в котором определение, следует ли представить фактографическое описание как факт на основе идентифицированного лингвистического элемента содержит

применение правил исключения в отношении лингвистических элементов фактографических описаний для удаления определенных фактографических описаний из рассмотрения,

оценивание фактографических описаний,

сравнение оценок каждого фактографического описания, оставшегося в рассмотрении, с порогом, и

для каждого фактографического описания, имеющего оценку, которая превышает порог, представление по меньшей мере части предложения, содержащего это фактографическое описание, как факта.

3. Способ по п.2, дополнительно содержащий обеспечение слов фактографических описаний признаком их частей речи.

4. Способ по п.3, в котором обеспечение слов фактографических описаний признаком их частей речи содержит применение признака существительного, когда слово может быть или глаголом или существительным.

5. Способ по п.4, в котором применение правил исключения содержит применение первого набора правил для синтаксических словосочетаний, которые играют роль подлежащих, и применение второго набора правил для синтаксических словосочетаний, которые играют роль дополнений.

6. Способ по п.5, в котором применение первого набора правил содержит исключение именных словосочетаний, имеющих модификатор подлежащих или дополнений «необъективный» или «мнение».

7. Способ по п.5, в котором применение второго набора правил содержит исключение именных словосочетаний-подлежащих, которые содержат существительные в определенном состоянии, не являющиеся именами собственными, исключение именных словосочетаний, которые содержат местоимения, и исключение именных словосочетаний-подлежащих, которые не появляются в начале текста.

8. Способ по п.5, дополнительно содержащий применение третьего набора правил без учета роли именного словосочетания.

9. Способ по п.8, в котором применение третьего набора правил содержит исключение фактографических описаний, в которых в пунктуации предложения присутствует вопросительный знак, и исключение предложений со словосочетаниями, которые включают в себя стоп-слово.

10. Способ по п.2, в котором оценивание фактографических описаний содержит оценивание только тех фактографических описаний, которые остались в рассмотрении после применения правил исключения или которые остаются в рассмотрении во время их применения.

11. Машиночитаемый носитель информации, содержащий команды, которые при выполнении процессором побуждают процессор выполнять действия, содержащие прием термина поиска, содержащего существительное, обнаружение релевантных электронных ресурсов, которые соответствуют терму поиска,

отображение списка релевантных электронных ресурсов и фрагментов релевантных

электронных ресурсов в списке, содержащих слова, соответствующие терму поиска, синтаксический анализ множества релевантных электронных документов для обнаружения фактографических описаний из предложений, которые содержат существительные из терма поиска и один или более глаголов, соответствующих

словам таблицы слов факта, построенной так, что она включает в себя список глаголов, определенных как указывающие на выражение факта, удаление из обработки для извлечения фактов тех частей релевантного электронного документа, которые содержат слова, не соответствующие терму поиска, и слова таблицы слов факта,

исследование обнаруженных фактографических описаний для идентификации лингвистических элементов фактографических описаний после удаления частей релевантных электронных документов,

определение, следует ли представить фактографическое описание как факт, релевантный терму поиска, на основе идентифицированного лингвистического элемента, путем применения правил исключения к фактографическим описаниям-кандидатам в отношении лингвистических элементов, оценивание фактографических описаний-кандидатов на основании достоверности соответствующего слова факта и на основании отдельных весов именных словосочетаний-подлежащих и дополнений и удаление фактографических описаний-кандидатов из рассмотрения в соответствии с правилами исключения и оцениванием фактографических описаний, и

представление по меньшей мере части предложения, которое содержит термы поиска и фактографическое описание, определенное как факт, релевантный терму поиска.

12. Машиночитаемый носитель информации по п.11, в котором действия также содержат получение множества документов при поиске в совокупности электронных документов для обнаружения тех документов, которые содержат терм поиска, причем поиск в этой совокупности выполняется для обнаружения тех документов, которые содержат терм поиска, перед синтаксическим анализом множества электронных документов.

13. Машиночитаемый носитель информации по п.11, в котором действия также содержат получение электронных документов и представление фактографических описаний перед приемом терма поиска и поиском в электронных документах и фактографических описаниях для обнаружения тех электронных документов и соответствующих фактографических описаний, которые относятся к терму поиска.

14. Машиночитаемый носитель информации по п.11, в котором действия дополнительно содержат сравнение оценки каждого фактографического описания, оставшегося в рассмотрении, с порогом, и

представление по меньшей мере части предложения, содержащего фактографическое описание, как факта, относящегося к терму поиска, для каждого фактографического описания, которое взято из электронного документа, содержащего терм поиска и имеющего оценку, которая превышает порог.

15. Машиночитаемый носитель информации по п.14, в котором оценивание фактографических описаний содержит оценивание только тех фактографических описаний, которые остаются в рассмотрении после применения правил исключения.

16. Компьютерная система, содержащая:

запоминающее устройство, содержащее множество электронных ресурсов, которые содержат текстовую информацию,

процессор, который принимает терм поиска, содержащий существительное,

обнаруживает релевантные электронные ресурсы, которые соответствуют терму поиска, отображает список релевантных электронных ресурсов и фрагментов релевантных электронных ресурсов в списке, содержащих слова, соответствующие терму поиска, и принимает запрос для представления фактов, которые относятся к терму поиска, из набора релевантных электронных документов,

причем упомянутый процессор выполняет синтаксический анализ релевантных электронных документов для обнаружения фактографических описаний из предложений, которые содержат существительное из термина поиска и один или более глаголов, соответствующих словам таблицы слов факта, построенной так, что она включает в себя список глаголов, определенных как указывающие на выражение факта, процессор удаляет из обработки для извлечения фактов те части релевантного электронного документа, которые содержат слова, не соответствующие терму поиска, и слова таблицы слов факта, процессор исследует обнаруженные фактографические описания для идентификации лингвистических элементов фактографических описаний после удаления частей релевантных электронных документов, определяет, следует ли представить фактографическое описание как факт на основе идентифицированного лингвистического элемента, и представляет по меньшей мере часть предложений, которые содержат фактографические описания, в отношении которых определено, что их следует представить как факт, и которые относятся к терму поиска.

17. Компьютерная система по п.16, дополнительно содержащая дисплей, при этом процессор представляет по меньшей мере часть предложений путем вывода на экран дисплея по меньшей мере частей предложений.

18. Компьютерная система по п.16, дополнительно содержащая сетевой интерфейс, при этом процессор представляет по меньшей мере часть предложений путем вывода упомянутых частей на другой компьютер через сетевой интерфейс.

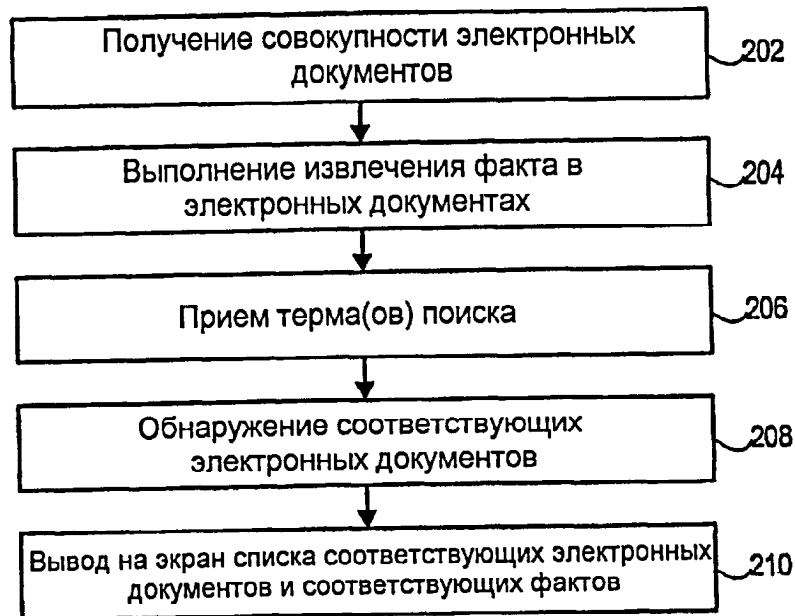
19. Компьютерная система по п.16, дополнительно содержащая сетевой интерфейс, при этом запоминающее устройство доступно процессору через сетевой интерфейс.

20. Компьютерная система по п.16, в которой процессор определяет, следует ли представить фактографическое описание как факт, путем

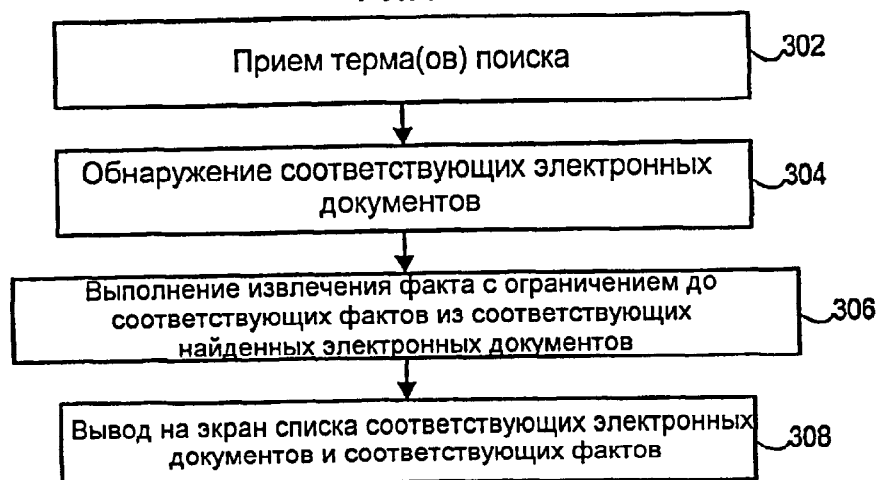
применения правил исключения в отношении лингвистических элементов фактографических описаний для удаления части фактографических описаний из рассмотрения,

оценивания фактографических описаний, сравнения оценки каждого фактографического описания, оставшегося в рассмотрении, с порогом, и

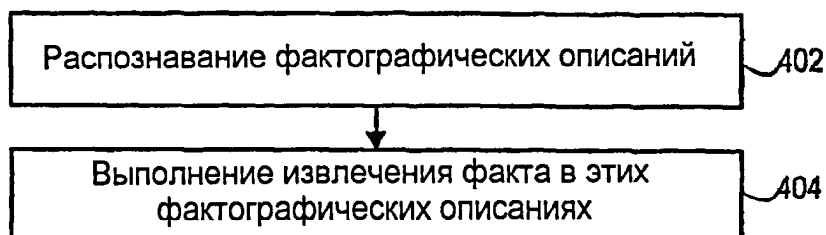
представления по меньшей мере части предложения, содержащего фактографическое описание, как факта, относящегося к терму поиска, для каждого фактографического описания, которое содержит терм поиска и имеет оценку, которая превышает порог.



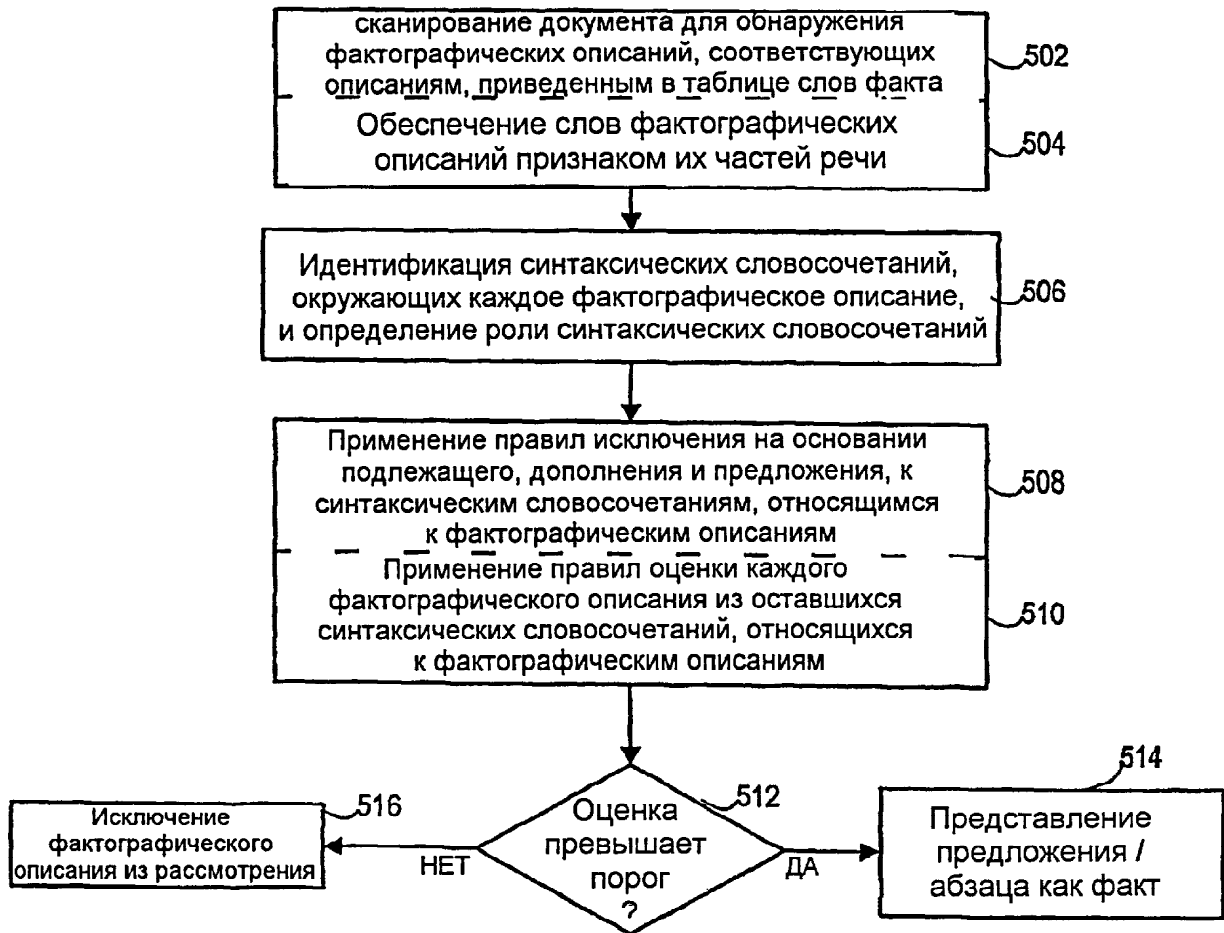
Фиг.2



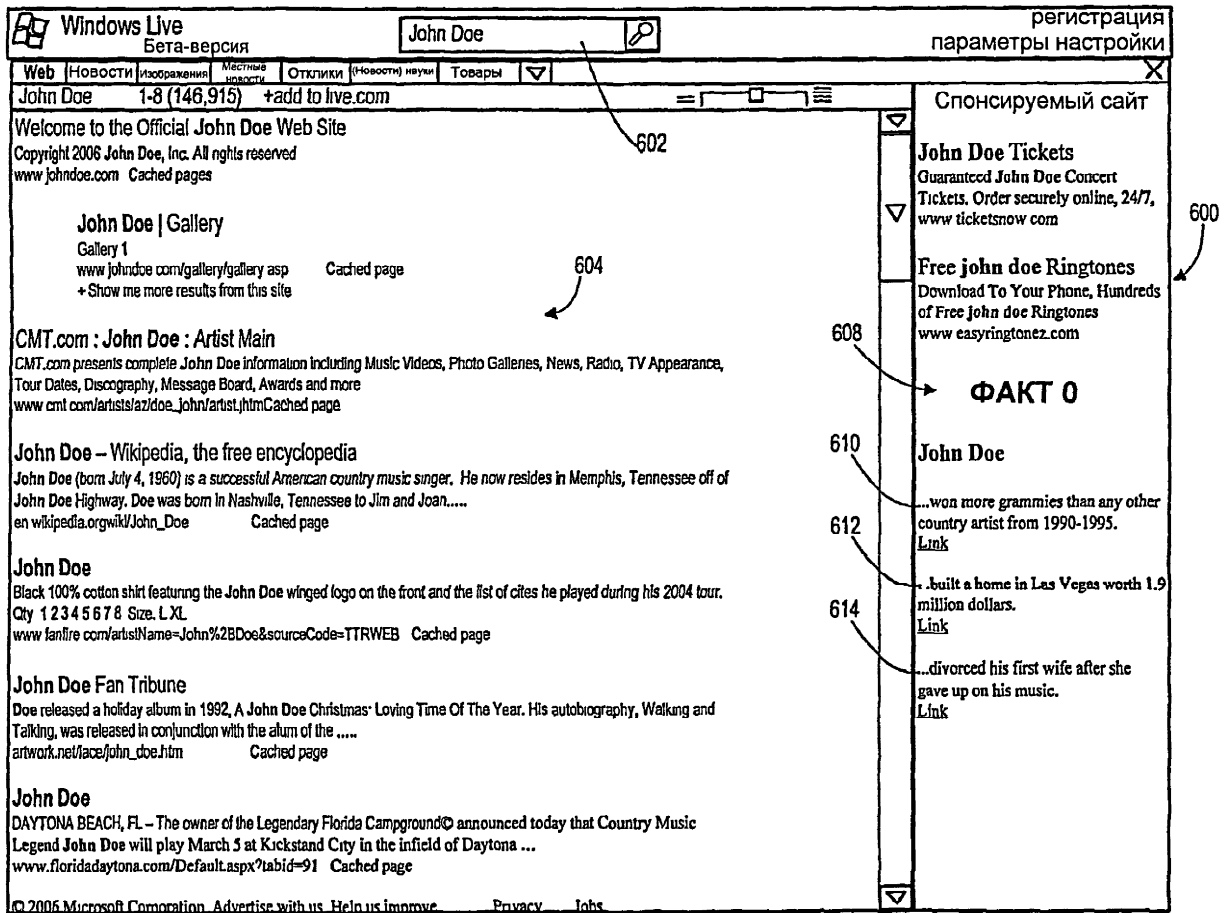
Фиг.3



Фиг.4



Фиг.5



Фиг.6