



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: 2009121432/08, 05.12.2007

(24) Дата начала отсчета срока действия патента:
05.12.2007

Приоритет(ы):

(30) Конвенционный приоритет:
05.12.2006 US 11/633,788

(43) Дата публикации заявки: 10.12.2010 Бюл. № 34

(45) Опубликовано: 10.08.2012 Бюл. № 22

(56) Список документов, цитированных в отчете о
поиске: US 2004/0006466 A1, 08.01.2004. US
2006/0206806 A1, 14.09.2006. US 2004/0030540
A1, 12.02.2004. US 2005/0125215 A1, 09.06.2005.
RU 2136038 C1, 27.08.1999. ROBB THOMAS.
Google as a Quick 'n Dirty Tool, TESL-EJ, vol.7,
No.2, сентябрь 2003, [онлайн],
[найдено 13.02.2012] в (см. прод.)(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 04.06.2009(86) Заявка РСТ:
US 2007/086560 (05.12.2007)(87) Публикация заявки РСТ:
WO 2008/070750 (12.06.2008)Адрес для переписки:
129090, Москва, ул. Б. Спасская, 25, стр.3,
ООО "Юридическая фирма Городисский и
Партнеры", пат.п.в. Ю.Д.Кузнецову,
рег.№ 595

(72) Автор(ы):

ГАО Цзяньфэнг (US),
ДОЛАН Уилльям Б. (US),
ХОНЬ Сяо-Уэнь (US),
ЧЖОУ Мин (US)

(73) Патентообладатель(и):

МАЙКРОСОФТ КОРПОРЕЙШН (US)

R U 2 4 5 9 1 C 2

R U 2 4 5 8 3 9 1 C 2

(54) ПРОВЕРКА ОШИБОК СОЧЕТАНИЙ СЛОВ НА БАЗЕ СЕТИ ИНТЕРНЕТ

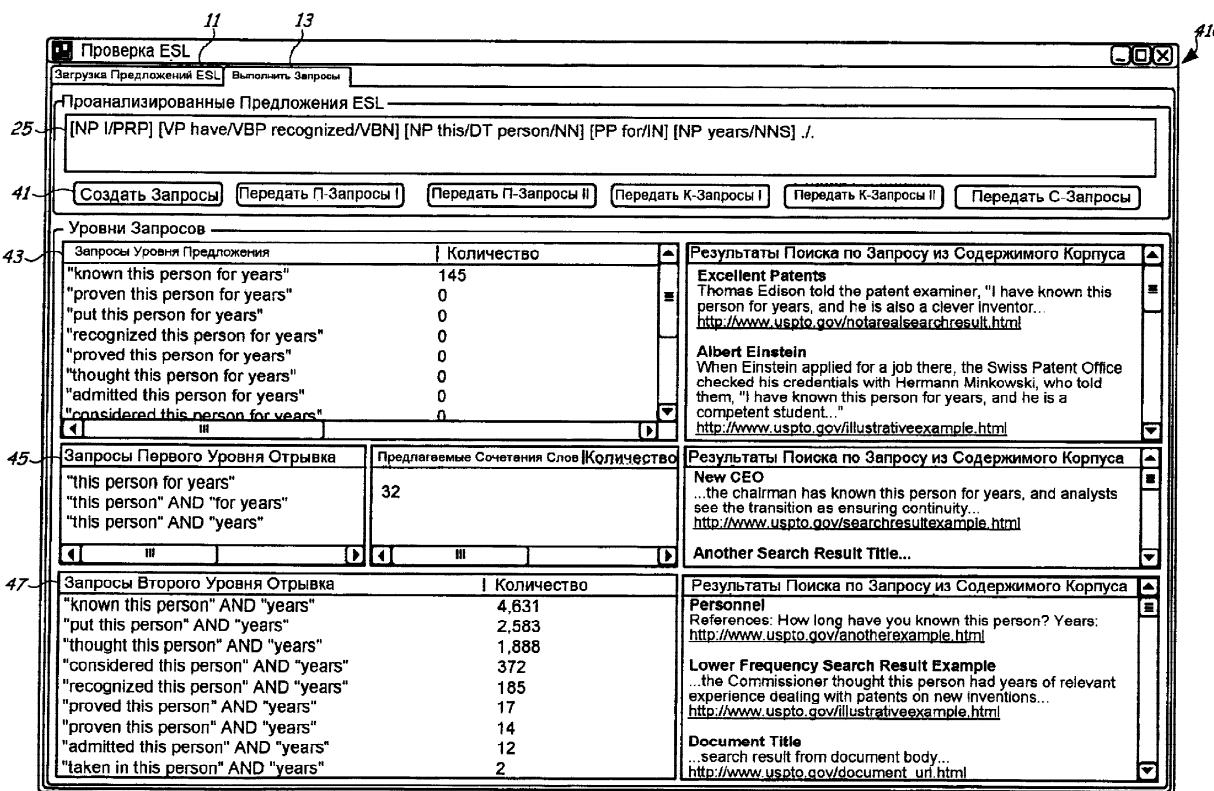
(57) Реферат:

Изобретение относится к средствам проверки сочетаний слов. Технический результат заключается в уменьшении ошибок при переводе сочетаний слов. Сравнивают одно или более сочетаний слов из образца текста с корпусом. Идентифицируют, являются ли упомянутые сочетания слов нестандартными в упомянутом корпусе. Через

устройство вывода предоставляют индикации о том, являются ли упомянутые сочетания слов нестандартными. Выполняют один или более поисков в Интернете, используя один или более терминов запроса, которые содержат каждое из одного или более упомянутых сочетаний слов. Составляют один или более терминов запроса с универсальным шаблоном, замещающим одно из слов в одном из

нестандартных сочетаний слов. Выполняют поиск эталона сочетаний слов для упомянутых терминов запроса. Идентифицируют результаты поиска с относительно большой долей замещения возможным словом

универсального шаблона. Представляют через устройство вывода результаты поиска с возможным словом в качестве потенциально правильных сочетаний слов. 3 н. и 14 з.п. ф-лы, 6 ил.



Фиг.4

(56) (продолжение):

Интернете:<URL:<http://www.cc.kyoto-su.ac.jp/information/testl-ej/ej26/int.html>>. C.-C.SHEI et al. An ESL Writer's Collocational Aid, Computer Assisted Language Learning, Vol.13, No 2, 2000, [онлайн], [найдено 13.02.2012] в Интернете:<URL: <http://www.corpus4u.org/forum/upload/forum/2005062509274575.pdf>>.

R U 2 4 5 8 3 9 1 C 2

RUSSIAN FEDERATION



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(19) RU (11) 2 458 391 (13) C2

(51) Int. Cl.
G06F 17/28 (2006.01)

(12) ABSTRACT OF INVENTION

(21)(22) Application: 2009121432/08, 05.12.2007

(24) Effective date for property rights:
05.12.2007

Priority:

(30) Convention priority:
05.12.2006 US 11/633,788

(43) Application published: 10.12.2010 Bull. 34

(45) Date of publication: 10.08.2012 Bull. 22

(85) Commencement of national phase: 04.06.2009

(86) PCT application:
US 2007/086560 (05.12.2007)

(87) PCT publication:
WO 2008/070750 (12.06.2008)

Mail address:

129090, Moskva, ul. B. Spasskaja, 25, str.3, OOO
"Juridicheskaja firma Gorodisskij i Partnery",
pat.pov. Ju.D.Kuznetsov, reg.№ 595

(72) Inventor(s):

GAO Tszjan'feng (US),
DOLAN Uill'jam B. (US),
KhON' Sjao-Uehn' (US),
ChZhOU Min (US)

(73) Proprietor(s):

MAJKROSOFT KORPOREJShN (US)

R U 2 4 5 8 3 9 1 C 2

(54) INTERNET-BASED COLLOCATION ERROR CHECKING

(57) Abstract:

FIELD: information technology.

SUBSTANCE: one or more collocations from a text sample are compared with the body. It is identified whether said collocations are non-standard in said body. An output device indicates whether said collocations are non-standard. One or more Internet searches are performed using one or more query terms which contain each of the one or more said collocations. One or more query terms are formed with a universal template which substitutes one of the words in one of the non-standard collocations. Search is performed for the collocation standard for said query terms. Search results are identified with relatively high ratio of substitution of a possible word of the universal template. Search results with

the possible word are presented through the output device as the potentially correct collocations.

EFFECT: reduced errors when translating collocations.

17 cl, 6 dwg

Запросы Узлового Глоссария	Количество	Результаты Поиска по Запросу из Словаря Корпуса
"known this person for years"	145	Excellent Patient The patient has been seen by her patient examiner, "I have known this person for years", and he is also a clever doctor.
"put me person for years"	0	
"thought this person for years"	0	
"brought this person for years"	0	
"proven this person for years"	0	
"thought this person for years"	0	
"admitted this person for years"	0	
"admitted this person for years"	0	

Запросы Второго Узлового Окна	Количество	Результаты Поиска по Запросу из Словаря Корпуса
"this person for years"	32	New CEO the chairman has known this person for years, and analyses their work.
"this person" AND "years"	1,886	Albert Einstein When Einstein was offered a job there, the Swiss Patent Office (inventor of the theory of relativity) asked him a question, who told them, "I have known this person for years, and he is a nice man." http://www.uspto.gov/tutorials/example.html
"this person" AND "years"	312	Another Search Result...
"this person" AND "years"	165	
"proven this person" AND "years"	17	
"thought this person" AND "years"	14	
"admitted this person" AND "years"	12	
"taken in this person" AND "years"	2	

Фиг.4

Уровень техники

При переводе с одного языка на другой часто возникают сложности в случаях, где значение отдельных слов дополнительно зависит от контекста, в котором они используются. Когда люди изучают второй язык, который не является для них родным, для них доступно множество выборов слов, перевод которых кажется правильным, но которые в действительности не используются носителями языка в определенных группах или сочетаниях слов. Подобные сочетания слов могут быть синтаксически или грамматически корректны в абстрактном смысле, но их применение ненормально, и зачастую для носителей языка они кажутся необычными или странными. Комбинации слов на некотором языке часто соответствуют установленным шаблонам, которые формируют единые лексические элементы, так что группы слов, которые имеют схожее значение и расположены в схожем грамматическом порядке, могут не входить в категорию какого-либо лексического элемента в рамках применения языка его носителями, независимо от того, является ли данный лексический элемент для носителей языка устоявшимся и ограничивающим использование альтернативных вариантов.

Например, для лица, изучающего английский в качестве второго языка, слова "strong" (крепкий) и "powerful" (сильный) могут рассматриваться как два альтернативных перевода одного слова с его родного языка. Еще одним примером являются слова "to make" (разрабатывать) и "to do" (делать). Без достаточного опыта в использовании английского языка учащийся может написать "I had a cup of powerful tea" (Я выпил чашку сильного чая) вместо "I had a cup of strong tea" (Я выпил чашку крепкого чая) или "I did a plan" (Я сделал план) вместо "I made a plan" (Я разработал план). Подобные ошибки сочетания слов могут быть типичны для носителя любого языка, который изучает любой другой язык, который не является для него родным. Таких ошибок, как правило, удается избежать только после достаточно долгого периода контекстуального опыта и интерактивной практики языка, причем, как правило, эти ошибки не могут быть разрешены с помощью словаря или какого-либо другого типа быстро и легкодоступного языкового справочника.

Настоящий раздел приведен исключительно для предоставления общего уровня техники и не предназначен для использования в качестве помощи при определении объема формулы изобретения.

Раскрытие изобретения

В различных вариантах осуществления, которые включают в себя способы, вычислительные системы и программное обеспечение, которое кодирует выполняемые инструкции для вычислительных систем, ошибки сочетаний слов могут быть автоматически проверены посредством локальных и сетевых корпусов (совокупностей текстов), таких как сеть Интернет. Например, согласно одному иллюстративному способу одно или более сочетаний слов из образца текста сравнивают с корпусом, таким как содержимое сети Интернет. Сочетания слов идентифицируют для определения того, являются ли они нестандартными в корпусе. Через устройство вывода предоставляют индикации о том, являются ли упомянутые сочетания слов нестандартными в корпусе. Далее могут быть предприняты дополнительные шаги, такие как поиск и предоставление через устройство вывода потенциально правильных сочетаний слов.

Раздел "Раскрытие изобретения" приведен, чтобы представить в упрощенной форме выборку концепций, которые подробно описываются ниже, в разделе "Осуществление изобретения". Раздел "Раскрытие изобретения" не предназначен ни для определения

5 ключевых или существенных отличительных признаков сущности формулы изобретения, ни для использования в качестве вспомогательного средства при определении объема формулы изобретения. Объем формулы изобретения не ограничивается реализациами, которые устраниют какие-либо или все недостатки, перечисленные в разделе "Уровень техники".

Краткое описание чертежей

10 Фиг.1 - иллюстрация пользовательского интерфейса для вычислительной системы, реализующей способ согласно иллюстративному варианту осуществления настоящего изобретения;

15 Фиг.2 - иллюстрация схемы последовательности операций способа согласно иллюстративному варианту осуществления настоящего изобретения;

20 Фиг.3 - иллюстрация схемы последовательности операций способа согласно иллюстративному варианту осуществления настоящего изобретения;

25 Фиг.4 - иллюстрация пользовательского интерфейса для вычислительной системы, реализующей способ согласно иллюстративному варианту осуществления настоящего изобретения;

30 Фиг.5 - структурная схема вычислительной среды, в которой могут быть реализованы некоторые варианты осуществления настоящего изобретения;

35 Фиг.6 - структурная схема еще одной вычислительной среды, в которой могут быть реализованы некоторые варианты осуществления настоящего изобретения.

Осуществление изобретения

40 На Фиг.1 проиллюстрирован интерфейс 10 проверки ошибок сочетаний слов для вычислительной системы, которая реализует способ для проверки ошибок сочетаний слов согласно иллюстративному варианту осуществления настоящего изобретения.

45 Интерфейс 10 проверки ошибок сочетаний слов может рассматриваться как пример для вариантов осуществления, которые включают в себя вычислительные системы, выполняемые инструкции, сконфигурированные для выполнения вычислительными системами, а также контексты, которые реализуют варианты осуществления способов. В нижесложенном описании приведены дополнительные детали различных иллюстративных вариантов осуществления настоящего изобретения. Наряду с тем, что в этот и в следующие чертежи включены определенные иллюстративные

50 компоновки и метки, а также примеры текста, которые сравниваются с помощью реализуемого компьютером способа, они предназначены для иллюстрации всего многообразия и широкого значения, предоставленного в настоящем описании и формуле изобретения.

55 В иллюстративном варианте осуществления на Фиг.1 интерфейс 10 проверки ошибок сочетаний слов образует панель графического отображения, связанную с программным приложением, то есть одну из множества панелей графического отображения, которые могут быть открыты одновременно в графическом пользовательском интерфейсе на устройстве вывода, таком как монитор компьютера.

60 Для одного иллюстративного варианта осуществления, направленного на проверку образцов текста, вводимого изучающими английский как второй язык (English as a Second Language, ESL), интерфейс 10 проверки ошибок сочетаний слов имеет метку "Проверка ESL" в строке 12 заголовка. ESL является идеальным приложением для одного иллюстративного варианта осуществления проверки ошибок сочетания слов, поскольку неправильные сочетания слов представляют большую часть ошибок студентов среднего уровня, изучающих английский как второй язык.

Следует понимать, что наряду с тем, что в нижесложенном описании многократно

цитируется иллюстративный вариант осуществления приложения проверки ESL, он представляет всего лишь иллюстративный пример, который показывает более широкие принципы, которые могут быть применены к множеству других вариантов осуществления. Например, ниже описаны другие варианты осуществления, которые могут действовать без интерфейса 10 проверки ошибок сочетаний слов, изображенного на Фиг.1, а другие дополнительные варианты осуществления могут быть направлены на учеников любого другого языка как второго языка, и другие варианты осуществления могут быть направлены, например, на детей, изучающих свой родной язык.

В этом конкретном варианте осуществления в верхней части интерфейса 10 проверки ошибок сочетаний слов присутствуют две вкладки, соответствующие двум различным режимам работы: вкладка 11 "Загрузить Предложения ESL" и вкладка 13 "Выполнить Запросы". На Фиг.1 проиллюстрирован пользовательский интерфейс 10 при выбранной вкладке 11 "Загрузить Предложения ESL" с отображением соответствующих кнопок, заголовков, текстовых окон и других элементов пользовательского интерфейса (или "виджетов"). На Фиг.4 проиллюстрирован интерфейс 10 проверки ошибок сочетаний слов при выбранной вкладке 13 "Выполнить Запросы" с отображением соответствующих элементов пользовательского интерфейса, описание которых следует ниже.

При выбранной вкладке 11 "Загрузить Предложения ESL" панель пользовательского интерфейса включает в себя множество различных интерактивных элементов пользовательского интерфейса ввода и вывода, описание которых приведено далее. Они разделены на две группы, входящие в верхнюю секцию с заголовком "Анализ Предложений" и нижнюю секцию с заголовком "Анализ Сочетаний Слов", соответственно.

В верхней части секции "Анализ Предложений" по горизонтали расположен ряд элементов интерфейса: кнопка 15 "Загрузить Из Проанализированных Файлов", кнопка 17 "Загрузить Из Необработанных Файлов", заголовок "Тип Словосочетания" перед комбинированным окном 19, в котором на данной фигуре выведена опция "Сочетание слов Глагол-Существительное" (описанное ниже), а также кнопка 21 "Извлечь Сочетания Слов". Под этими элементами управления расположено текстовое окно 23 с горизонтальной и вертикальной полосами 22, 24 прокрутки, а также строка 20 заголовка с надписью "Исходное Предложение". Ниже расположено текстовое окно 25 с заголовком "Проанализированное Предложение". В целом, эти элементы интерфейса предоставляют возможность пользователю загружать предложения из образца текста и подготавливать их для сравнения в целях выявления ошибок с сочетаниями слов в контексте, доступном через сеть.

Текстовое окно 23 включает в себя несколько образцов предложений из образца текста, такого как документ, написанный изучающим ESL. Исходя из сжатого размера бегунка 26 вертикальной полосы прокрутки 24 текстового окна 23, можно заметить, что отображаемые на фигуре предложения представляют лишь малую часть общего количества предложений, загруженных в текущий момент в текстовое окно, причем пользователь может открыть их путем перемещения бегунка 26. Отображенные в текстовом окне 23 предложения представляют типичные ошибки, которые могут допустить студенты среднего уровня, изучающие английский как второй язык: "I did a plan" (Я сделал план), "I have recognized this person for years" (Я узнал этого человека много лет) и "I had a cup of powerful tea" (Я выпил чашку сильного чая).

Эти предложения могут быть вручную загружены из документа, который открыт в

другом приложении или они могут быть автоматически выбраны как содержащие ошибки типа ESL и загружены из другого документа или множества документов, после того как пользователь инициирует соответствующий процесс путем выбора 5 кнопки 15 "Загрузить Из Проанализированных Файлов" или кнопки 17 "Загрузить Из Необработанных Файлов". В еще одном варианте осуществления некое приложение, такое как приложение обработки текста, может содержать триггер проверки ошибок ESL, который может быть активирован по умолчанию, благодаря чему в случае 10 детектирования ошибок типа ESL в тексте, вводимом в программу обработки текста, автоматически появляется панель с рекомендацией применения интерфейса 10 проверки ошибок сочетаний слов.

Наличие кнопок 15 и 17 предполагает, что когда в приложение вводятся образцы текста, они могут быть уже проанализированы или еще не проанализированы. Если образец текста еще не проанализирован, то он, например, может быть подвергнут 15 анализу до выполнения дальнейших этапов. Проанализированное предложение показано в текстовом окне 25. В частности, это проанализированная версия второго предложения, показанного в текстовом окне 23 и имеющего локальное выделение для индикации того, что оно выбрано пользователем, например, путем выполнения на нем 20 левого щелчка мышью или прикосновения к нему световым пером, или с помощью некоторого другого механизма пользовательского ввода.

В последнее время технологии анализа активно изучаются в области обработки языка и понимания естественного языка. Анализ может включать в себя такие задачи, как маркировка частями речи, разбиение на отрывки и семантическое маркирование. 25 В варианте осуществления, показанном на Фиг.1, предложение в текстовом окне 25 было подвергнуто маркировке частями речи и разбиению на отрывки. Другие варианты осуществления могут содержать другие комбинации задач анализа, включающие в себя некоторые комбинации, которые направлены на отличные от 30 западных языков, например, сегментирование слов.

В проанализированном предложении в текстовом окне 25 за каждым словом исходного предложения следует косая черта (то есть "прямой слэш"), отделяющая это слово от двух- или трехбуквенного тега части речи, присвоенного этому слову. Некоторые иллюстративные опции для тегов частей речи приведены ниже. За точкой в 35 конце предложения также следует косая черта, отделяющая ее от другой точки и выполняющая роль эквивалента тега части речи для обозначения функции пунктуации. Группы из одного или двух (или более - для других примеров) слов предложения также заключаются в скобки, соответствующие границам отрывка, причем сразу после открывающей скобки каждой пары скобок устанавливается 40 двухбуквенная метка типа отрывка. Теги части речи, представленные в этом предложении, включают в себя "PRP" для личного местоимения; "VBP" для настоящего времени глагола, отличного от единственного числа третьего лица; "VBN" для глагола с причастием прошедшего времени; "DT" для определяющего слова; "NN" 45 для имени нарицательного единственного числа или равнозначенного; "IN" для предлога или подчинительного союза; и "NNS" для имени нарицательного множественного числа. Иллюстративный список тегов частей речи согласно одному примеру осуществления приведен ниже, однако следует отметить, что в других вариантах 50 осуществления могут использоваться другие схемы маркировки частями речи с другими определениями и большей или меньшей степенью детализации. Для настоящего примера осуществления некоторые теги проиллюстрированы с помощью примеров на английском языке.

Таблица 1
Иллюстративный список тегов частей речи

1. CC - Сочинительный союз (например, "and", "but", "nor", "or", "yet", "plus", "minus", "times", "over" (в смысле "разделить на"))
2. CD - Количествочислительное
- 5 3. DT - Определяющее слово (включает в себя артикли и неопределенные определяющие слова, например, "a", "an", "every", "no" (как артикль), "the"; и, например, "another", "any", "some", "each", "either", "neither", "that", "these", "this", "those", некоторые значения "all", "they")
4. EX - "there" в экзистенциальном значении
5. FW - Иностранные слова
6. IN - Предлог или подчинительный союз
7. JJ - Прилагательное
- 10 8. JJR - Прилагательное, сравнительная степень
9. JJS - Прилагательное, превосходная степень
10. LS - Маркер элемента списка
11. MD - Модальный глагол (например, "can", "could", "may", "might", "must", "shall", "should", "will", "would")
12. NN - Существительное, нарицательное, единственное число или равнозначное
13. NNS - Существительное, нарицательное, множественное число
14. NNP - Имя собственное, единственное число
- 15 15. NNPS - Имя собственное, множественное число
16. PDT - Предeterminator
17. POS - Притяжательное окончание
18. PRP - Местоимение, личное
19. PRP\$ - Притяжательное местоимение (например, "'s", "s'", "")
20. RB - Наречие
- 20 21. RBR - Наречие, сравнительная степень
22. RBS - Наречие, превосходная степень
23. RP - Служебное слово
24. SYM - Символ
25. TO - Частица "To"
26. UH - Восклицание
27. VB - Глагол, базовая форма
28. VBD - Глагол, прошедшее время
29. VBG - Глагол, герундий или причастие настоящего времени
30. VBN - Глагол, причастие прошедшего времени
31. VBP - Глагол, настоящее время, отличный от третьего лица единственного числа
32. VBZ - Глагол, настоящее время, третье лицо единственное число
33. WDT - Вопросительное определяющее слово
34. WP - Вопросительное местоимение
- 30 35. WP\$ - Притяжательное вопросительное местоимение ("whose")
36. WRB - Вопросительное наречие

Как показано проанализированным предложением в текстовом окне 25, проанализированный образец текста также был разбит на отрывки, причем каждый определенный отрывок заключен в пару скобок и тип отрывка указан с помощью метки перед содержимым каждой пары скобок. Разбиение на отрывки является относительно легкой задачей анализа, которая может быть выполнена на основании поверхностной и локальной информации. При разбиении на отрывки предложения разделяются на неперекрывающиеся сегменты таким образом, чтобы каждый отрывок содержал одно главное слово и связанные с ним другие слова. Так, в настоящем иллюстративном варианте осуществления разбиение на отрывки может рассматриваться как разделение предложения на группы, в частности, на наименьшие групповые блоки, которые могут быть определены, чтобы предотвратить наложение групп. Разбиение на отрывки предоставляет возможность идентификации синтаксической структуры текста, а также отношений или зависимостей между группами. Например, одна именная группа может быть подлежащим глагольной группы, а вторая именная группа может быть дополнением этой глагольной группы.

Иллюстративные отрывки на Фиг.1 являются хорошими примерами этого. Первый отрывок снабжен меткой NP для именной группы и основан на единственном главном слове "I" - подлежащем этого предложения. Второй отрывок снабжен меткой VP для глагольной группы и включает в себя слова "have" и "recognized": два отдельных глагола совместно формируют причастие прошедшего времени глагола "to recognize".

Третий отрывок также снабжен меткой NP и включает в себя главное слово "person" вместе со связанным, поддерживающим словом, чтобы образовать единую концепцию "this person" в качестве дополнения этого предложения. Четвертый отрывок снабжен меткой PP для предложной группы, а пятый отрывок снабжен меткой NP для еще одной именной группы, каждая из которых содержит одно слово. Точка, снабженная тегом точки, не входит в отрывки.

Система анализа, связанная с интерфейсом 10 проверки ошибок сочетаний слов или используемая им, может быть подвергнута обучению, чтобы автоматически, точно и надежно выполнять маркировку частями речи и разбиение на отрывки согласно способам, которые хорошо знакомы специалистам в области обработки естественных языков.

Нижняя секция панели пользовательского интерфейса с заголовком "Анализ Сочетаний Слов" включает в себя ряд текстовых окон 27, 29, 31 и 33, каждое из которых имеет строку заголовка и, по меньшей мере, одну полосу прокрутки. Стока заголовка текстового окна 27 имеют надпись "Сочетания Слов ESL", и показанное текстовое окно включает в себя два отрывка из предложения в текстовом блоке 25: глагольную группу "have recognized" и именную группу "this person" - дополнение этого предложения. Вышеупомянутые группы показаны с тегами частей речи, метками отрывков и скобками в следующей форме: "[VP have/VBP recognized/VBN] [NP this/DT person/NM]". Это сочетание слов было предоставлено в текстовом окне 27 из-за того, что комбинированное окно 19 типа сочетания слов установлено в значение "Сочетание слов Глагол-Существительное", то есть была выбрана опция, указывающая сочетания слов с глагольной группой и следующей именной группой, которые расположены рядом друг с другом. Сочетание слов, соответствующее этой установке, было извлечено из предложения в текстовое окно 25.

Текстовое окно 29 содержит возможные сочетания слов, что отражено в его строке заголовка. Возможные сочетания слов содержат ту же именную группу "this person", что и сочетание слов в текстовом окне 27, но уже с различными другими глаголами прошедшего времени, которые предшествуют этой именной группе. Эти возможные глаголы представляют собой возможные замены глагола "recognized" в качестве главного глагольного слова в прошедшем времени после слова "have" в глагольной группе. Они включают в себя глаголы "known", "admitted", "thought" и т.д. Среди этих возможных сочетаний нет исходного сочетания слов, поскольку сравнение исходного сочетания слов с содержимым корпуса показало, что оно либо редко используется, либо отсутствует в содержимом данного корпуса, и, следовательно, оно является нестандартным и, вероятно, неправильным.

Интерфейс 10 проверки ошибок сочетаний слов или приложение, связанное с ним или используемое с ним, может, следовательно, индицировать, что исходное сочетание слов является нестандартным в содержимом корпуса, и, следовательно, есть вероятность, что оно неправильное. В иллюстративном варианте осуществления с Фиг.1 эта индикация может быть предоставлена пользователю, например, в текстовом окне 33. В текстовом окне 33 указывается статус сочетания слов, что отражено в строке заголовка этого окна. В упомянутом окне указывается, что это сочетание слов является нестандартным в содержимом корпуса, и есть вероятность его неправильности, и под главным глагольным словом в сочетании слов, которое было определено как нестандартное, выводится волнистая линия 32, указывающая слово, которое является кандидатом на замену. Пользователь может выбрать одно из возможных сочетаний слов в текстовом окне 29, например, путем двойного щелчка на

желаемом сочетании слов, чтобы заменить сочетание слов, определенное как нестандартное. Эта индикация также может быть предоставлена, например, в отдельной программе обработки текста, программе для навигации по сети или другом приложении путем добавления метки к нестандартному сочетанию слов. Для привлечения внимания пользователя подобная метка может иметь форму выделения или подчеркивания нестандартного или неправильного сочетания слов, например, посредством волнистой линии яркого цвета.

Возможные сочетания слов в текстовом окне 29 предоставлены как результат

10 сравнения сочетания слов из текстового окна 27 с содержимым, доступным в большом корпусе. В этом иллюстративном варианте осуществления корпус может включать в себя локальный корпус и/или сетевой корпус с содержимым, которое хранится на распределенных ресурсах и которое доступно через сеть, такую как сеть Интернет. В 15 других вариантах осуществления корпус может включать в себя содержимое других доступных сетей, таких как интранет, глобальная сеть, локальная сеть или некоторый другой тип сети. Содержимое, полученное путем поиска в сети Интернет, где в качестве корпуса используется сеть Интернет, может быть использовано вместе с поиском в обычном корпусе, таком как корпус Wall Street Journal, например.

20 Присутствию сочетания слов в обычном корпусе может быть присвоен определенный весовой коэффициент, который больше весового коэффициента присутствия этого сочетания слов в содержимом сети Интернет, поскольку в корпусе правильное применение сочетаний слов обеспечивается в большей степени, чем в содержимом сети Интернет, где может присутствовать множество различных источников и включений 25 случайной информации. Тем не менее, содержимое сети Интернет также обеспечивает значительные преимущества, обусловленные его гораздо большим объемом относительно любого обычного корпуса. Присутствие в обычном корпусе правильного сочетания слов, соответствующего любому заданному сочетанию слов в 30 образце текста, маловероятно. С другой стороны, было выявлено, что, по крайней мере, на английском языке в сети Интернет может быть найдено практически любое сочетание слов, использование которого правильно, или, по меньшей мере, может быть обнаружено структурно эквивалентное сочетание слов путем поиска различных терминов на основании заданного сочетания слов, как описано ниже. Также было 35 выявлено, что правильное сочетание слов для замены может быть обнаружено в сети Интернет для, по меньшей мере, наиболее неправильных сочетаний слов, используемых в типовых образцах текста, например, изучающими английский как второй язык. Если поиск конкретного предложения или сочетания слов в сети 40 Интернет завершается неуспешно, то это является основанием для высокой степени уверенности в неправильности этого предложения или сочетания слов. Результаты поиска могут быть оценены, чтобы определить, встречаются ли схожие альтернативные варианты с гораздо большей частотой, чем заданное сочетание слов, и на основании этого может быть определена неправильность этого сочетания.

45 На момент написания настоящего документа большая часть текстового содержимого, доступного в сети Интернет, составлена на английском языке. Содержимое на множестве существующих в настоящее время языков представлено в Интернете в небольших объемах, и даже языки, разговорное применение которых 50 широко распространено, имеют относительно скромную долю в содержимом сети Интернет. Хинди, например, является третьим разговорным языком по распространенности в мире, но, по крайней мере, в одном исследовании было выявлено, что он представлен в содержимом Интернета в меньших объемах, чем

исландский язык, количество разговаривающих на котором меньше одной тысячной разговаривающих на хинди. Эффективность проверки ошибок сочетаний слов частично зависит от размера образца доступного корпуса, как базы сравнения для сочетаний слов в образце текста, который должен быть проверен. Следовательно, для проверки ошибок сочетаний слов на основе сети Интернет на других языках могут быть использованы специализированные способы, которые могут не требоваться при проверке образца текста на английском.

Эти способы могут включать в себя, например, конкретный поиск или отбор 10 содержимого на целевом языке, который совпадает с языком образца текста, или применение специализированной сети, корпуса или цифровой библиотеки, в дополнение к сети Интернет или содержимому, доступному через сеть. Эти способы также могут включать в себя использование поисковой утилиты для индексирования и поиска содержимого, кодирование которого отличается от стандарта Unicode, 15 например, или использование специализированных схем кодирования символов или форматов трансформации для возможности чтения, индексирования и поиска в нестандартных форматах кодирования. Присутствуют признаки того, что большой объем содержимого на языках, которые используют отличные от латинского 20 алфавита системы правописания, даже подавляющая часть содержимого сети Интернет для некоторых языков, существует в нестандартных форматах кодирования, так что поиск с возможностью индексирования и поиска этих нестандартных 25 форматов кодирования может привести к гораздо лучшим результатам. Согласно различным вариантам осуществления подобные способы могут улучшить эффективность системы проверки сочетаний слов на основе сети Интернет при ее применении для наиболее широкого возможного набора языков.

Дополнительные стратегии для ограничения или выбора содержимого сети 30 Интернет, в котором выполняется поиск, могут включать в себя применение предпочтений категорий или классификаций URL-адресов. Некоторые URL-адреса могут быть специальным образом каталогизированы для предпочтительного поиска и/или для применения более высокого весового коэффициента к сочетаниям слов, обнаруживаемым в их содержимом, если известно, что их содержимое включает в себя 35 большие библиотеки и/или корпусы надежного правильного использования языка. Это также может включать в себя предпочтительное сужение поиска или взвешивание 40 результатов поиска на основании доменов верхнего уровня. Например, добавление порядка поиска или взвешивания результатов может быть применено к URL-адресам с доменами верхнего уровня "edu" или "gov", в противоположность доменам верхнего уровня "com" или "org". Для поисков, используемых в проверке сочетаний слов для 45 отличных от английского языков, поиски содержимого сети Интернет на целевом языке также могут включать в себя сужение или ограничение поиска или применение весового коэффициента к его результатам на основании доменов верхнего уровня с кодами стран, назначенных странам, в которых целевой язык является официальным или используется в значительной мере. Например, приложение проверки сочетаний 50 слов, нацеленное на проверку образца текста на французском языке, может сузить или ограничить поиск содержимого сети Интернет только сайтами с доменами верхнего уровня "fr", "be", "ch" или "ca" (коды Франции, Бельгии, Швейцарии и Канады, соответственно) или другими доменами верхнего уровня с кодами стран, где французский является официальным языком или используется в значительной мере.

В текстовом окне 31 показана часть длинного списка всех доступных глаголов, которые были обнаружены в контексте корпуса, где в сочетании слов главное слово

глагольной группы предшествует словам "this person". Исходя из размера и положения 5
бегунка вертикальной полосы прокрутки на правой стороне текстового окна 31
можно заметить, что список гораздо больше, чем часть доступных глаголов,
выводимых в текущий момент в текстовом окне 31 на Фиг.1. Большинство доступных
10 сочетаний, приведенных в текстовом окне 31, не были выбраны в качестве возможных
сочетаний слов как в списке в текстовом окне 29, поскольку они не набрали
достаточно высокой оценки по схеме оценки нечеткого совпадения на основании
15 набора критериев, по которым определяется правильность совпадения исходного
сочетания слов и его замена на сочетание слов, которое дает исходное значение в
правильной форме.

Иллюстративные способы для сравнения сочетания слов в текстовом окне 27, для 15
идентификации того, что оно является нестандартным в содержимом корпуса, и для
предоставления предлагаемых замен для него подробно описаны со ссылкой на
остальные фигуры.

Фиг.2 представляет собой иллюстрацию схемы последовательности операций 20
способа 50 проверки сочетаний слов согласно иллюстративному варианту
осуществления, определенные аспекты которого аналогичны аспектам интерфейса 10
25 проверки ошибок сочетаний слов, изображенного на Фиг.1. Основные этапы
способа 50 показаны в сплошных контурах на схеме последовательности операций с
Фиг.2. Упомянутые основные этапы включают в себя этап 51, на котором сравнивают
одно или более сочетаний слов из образца текста с корпусом, как например, путем
передачи содержимого, включающего в себя упомянутое сочетание слов, в виде
30 поискового запроса в сети Интернет; этап 53, на котором идентифицируют, являются ли
упомянутые сочетания слов нестандартными в данном корпусе; и этап 55, на
котором через устройство вывода предоставляют индикации о том, являются ли
упомянутые сочетания слов нестандартными. Как показано посредством рекурсивной
35 стрелки, этапы 51 и 53 могут повторяться друг за другом, чтобы выполнять
различные типы сравнения сочетаний слов из образца текста с корпусом и чтобы
идентифицировать, являются ли эти сочетания слов нестандартными в корпусе,
согласно результатам различных типов сравнений.

Этапы 51, 53 и 55 также могут включать в себя дополнительные этапы, такие как 40
показанные в пунктирных контурах, или дополнительные этапы могут
предшествовать или следовать за основными этапами. Эти дополнительные этапы
включают в себя, например, этап, на котором принимают индикацию или
идентифицируют сочетания слов в тексте, как например, путем приема
пользовательского ввода или выбора или другой формы индикации образца текста,
как на этапе 61. Также в приложении обработки текста или некотором другом
45 программном приложении может быть включена функция по умолчанию, согласно
которой вводимый человеком текст отслеживается и сканируется на предмет наличия
неправильных сочетаний слов или других признаков применения неродного языка, и
при обнаружении подобных признаков запускаются дополнительные этапы, как на
этапе 63. Образец текста также может быть проанализирован, как для случая этапа 65
и как иллюстративно описано со ссылкой на Фиг.1, хотя к образцам текста, которые
уже были подвергнуты анализу, например, могут быть применены другие варианты
50 осуществления настоящего способа.

При сравнении с корпусом могут быть выбраны конкретные типы сочетаний слов, 55
как на этапе 67. На этом этапе для сравнения с корпусом и выполнения
дополнительных этапов проверки ошибок сочетаний слов выбирают сочетания слов,

такие как сочетания слов глагол-существительное, сочетания слов предлог-существительное, сочетания слов прилагательное-существительное или сочетания слов глагол-наречие. В некоторых вариантах осуществления может использоваться какая-либо одна из этих отдельных категорий сочетаний слов, тогда как в других вариантах осуществления может использоваться любое количество или все эти типы сочетаний слов и/или дополнительные типы сочетаний слов. Было выявлено, что при использовании четырех типов сочетаний слов, перечисленных для этапа 67, охватывается существенная часть всех ошибок сочетаний слов, допускаемых

10 типичными пользователями неродного языка. При выборе этих конкретных типов сочетаний сначала идентифицируют части речи в образце текста, как например, путем этапа подготовительного анализа, как на этапе 65 и как описано со ссылкой на Фиг.1.

Этап 53, на котором идентифицируют, являются ли сочетания слов нестандартными в корпусе, может выполняться многократно для множества отдельных запросов, 15 которые основаны на сочетании слов в различных форматах, в результате чего формируется окончательное определение, основанное на всех использованных запросах, согласно которому определяется, являются ли сравниваемые с корпусом сочетания нестандартными в контексте этого корпуса.

20 Один иллюстративный вариант осуществления дополнительно проиллюстрирован в аналогичных этапах 351 и 353 на Фиг.3. В этом варианте осуществления сравнение сочетаний слов из образца текста с корпусом, как на этапе 51 на Фиг.2, может включать в себя стратегию нечеткого совпадения, согласно которой в корпусе выполняют поиск терминов запросов, которые включают в себя либо целое

25 предложение, либо сокращенное предложение, либо пару отрывков, либо пару слов, которые образуют заданное сочетание, как показано на этапах 361, 363, 365 и 367, соответственно, на Фиг.3. Сокращенное предложение формируется путем удаления вспомогательных слов из предложения, которое включает в себя данное сочетание

30 слов, причем в иллюстративном варианте осуществления вспомогательные слова представляют собой слова из предложения, которые могут быть предварительно определены как необязательные для сохранения заданного сочетания слов в этом предложении. Например, для предложения "I have recognized this person for years" (Я знал этого человека много лет) шаблон запроса сокращенного предложения,

35 основанный на неправильном сочетании, может иметь форму "have recognized this person" (узнал этого человека). Термины запроса из пары отрывков и пары слов могут быть сформированы для конкретных слов или отрывков, содержащих это сочетание слов, либо из непосредственно смежных слов или отрывков, либо из слов или

40 отрывков, которые находятся в предварительно выбранном диапазоне близости относительно друг друга внутри одного предложения, как например, при наличии не более одного или двух других слов, которые разделяют их. При поиске с нечетким совпадением в результате поиска сокращенного предложения, пары отрывков и пары слов также могут быть получены результаты с другим, но эквивалентным порядком

45 слов или с замещающим словом для определенных функциональных слов или других типов слов, так что для нечеткого совпадения сочетание "recognize this person" (узнать этого человека) и сочетание "recognize that person" (узнать того человека) могут рассматриваться как эквивалентные.

50 Сочетание слов может быть идентифицировано как нестандартное, если термины запроса, которые содержат данное сочетание слов, не набирают оценку выше предопределенного порогового значения, указывающего значительное присутствие нечетких совпадений терминов запроса в корпусе. При идентификации того, являются

ли сочетания слов нестандартными в корпусе, как на этапе 53 на Фиг.2, могут быть выполнены различные типы поисков терминов запроса, включая поиск совпадения предложения или сокращенного предложения, как на этапах 371 и 373, соответственно, или может быть определено пороговое значение оценки для 5 совпадений, найденных для пар отрывков или пар слов, как на этапах 375 и 377, соответственно, на Фиг.3. Поскольку предложения и сокращенные предложения гораздо более специфичны, одно совпадение или небольшое количество совпадений с одним из них может рассматриваться как достаточное для индикации того, что 10 сочетание слов не является нестандартным или неправильным. Поскольку пары отрывков или пары слов менее ограничены и более обычны, вероятность их обнаружения в, по меньшей мере, нескольких результатах поиска более высока даже при их неправильности, поскольку относительно небольшая часть содержимого сети 15 Интернет написана неискусствами пользователями языка. Следовательно, для совпадения пары отрывков или пары слов может быть установлено более высокое пороговое значение, достаточное для индикации относительного преобладания указанной пары отрывков или слов, свидетельствующего о соответствии данного сочетания слова нормальному использованию, в отличие от относительно 20 незначительного представления в корпусе, свидетельствующего о том, что данное сочетание слов является нестандартным, согласно данному варианту осуществления.

Оценка совпадений может включать в себя основной счетчик, который присваивает определенную накопительную оценку при каждом совпадении пары отрывков или 25 пары слов, например, и который только указывает, что сочетание слов скорее является правильным, чем нестандартным, если обнаруживается достаточное количество совпадений, обеспечивающих достижение минимальной общей оценки. Оценки для результатов с различными запросами также могут быть комбинированы, причем больший вес придается более специфичным терминам запроса. Например, в одном 30 иллюстративном примере весовой коэффициент, присваиваемый при совпадении пары отрывков, в десять раз больше весового коэффициента, присваиваемого при совпадении пары слов. При определении статуса сочетания слов сокращенные предложения или полные предложения также могут быть включены в оценку совпадения, а не использоваться отдельно. Например, каждому совпадению, 35 обнаруживаемому для сокращенного предложения, которое содержит сочетание слов из образца текста, присваивается весовой коэффициент, который в двадцать раз больше соответствующего коэффициента для совпадения пары отрывков и в двести раз больше соответствующего коэффициента для совпадения пары слов. Независимо 40 от способа, используемого для оценки совпадений, процесс может завершиться на этапе 355, на котором предоставляют индикацию о том, что эти сочетания слов являются нестандартными, как на этапе 381, или что эти сочетания не являются нестандартными, как на этапе 383.

Присваивание весовых коэффициентов различным шаблонам запроса и пороговое 45 значение оценки совпадения, используемые для различия обычно используемых сочетаний слов от нестандартных сочетаний слов, могут быть откалиброваны или настроены для большей или меньшей чувствительности либо разработчиком программного обеспечения, либо конечным пользователем в рамках 50 пользовательских настроек. Пользователь может уменьшить упомянутое пороговое значение, чтобы получить более быстрый процесс, хотя и с более высокой вероятностью получения ошибочных результатов, или пользователь может повысить пороговое значение, возможно, потратив больше времени на поиск, но с большей

степенью устранения каких-либо ошибочных сочетаний из образца текста.

Поскольку порядок терминов запроса на Фиг.3 построен по принципу от более специфического к более общему и менее ограниченному, сбор результатов поиска при каждом последующем этапе, вероятно, займет больше времени, чем для предыдущего этапа. Несмотря на то, что все этапы возвращения результатов поиска, вероятно, потребуют не более доли секунды, производительность системы в целом может быть еще более повышена и общее время процесса может быть сокращено путем завершения этапов 351 и 353 сразу после получения достаточной индикации об обнаружении совпадения, без выполнения остающихся этапов, каждый следующий из которых требует все больше времени. Например, если для заданного предложения или сокращенного предложения обнаруживается множество совпадений, то можно будет избежать расходования относительно большей доли секунды, необходимой для выполнения поисков пары отрывков или пары слов. Поиски одного и того же сочетания слов посредством различных терминов запроса, следовательно, могут быть выполнены последовательно до тех пор, пока один из терминов запроса не предоставляет результаты поиска, которые удовлетворяют предварительно выбранному пороговому значению для совпадения сочетания слов, или пока все термины запроса, которые содержат это сочетание слов, не будут использованы без достижения предварительно выбранного порогового значения.

Возвращаясь к дополнительным этапам, показанным на Фиг.2, после предоставления индикаций о том, являются ли определенные сочетания слов нестандартными, согласно одному иллюстративному варианту осуществления, приложение проверки ошибок сочетаний слов может продолжить поиск и предоставить возможные замещающие слова, чтобы исправить неправильные сочетания слов, как показано на этапах 71, 73, 75 и 77. При этом могут быть составлены термины запроса с универсальным шаблоном вместо одного или более слов в одном или более нестандартных сочетаниях слов, как на этапе 71. Универсальный шаблон может выполнять роль свободного заполнителя, который вставляется в термин запроса вместо подозрительного слова, которое определяется как потенциально неправильное, причем заполнитель может соответствовать любому слову, вставленному в положение подозрительного слова в остальной части термина запроса. Универсальный шаблон может быть использован в поиске путем замены одного из слов символом, таким как знак звездочки, которая становится на место подозрительного неправильного слова.

Для терминов запроса может быть выполнен поиск эталона сочетания слов, что предоставляет потенциально возможные сочетания слов, как на этапе 73, такие как слова с элементами универсального шаблона. Этапоном сочетаний слов может быть сеть Интернет или другой сетевой или локальный корпус, причем данный эталон сочетаний слов также может включать в себя эталон из специализированного словаря сочетаний слов, получаемый, например, путем сканирования результатов поиска в сети Интернет для исключения сочетаний слов, не входящих в словарь сочетаний слов.

Поиск эталона сочетаний слов может дать результаты с сегментами текста, в которых вместо символа, представляющего универсальный шаблон заполнителя, используется какое-либо слово или группа. Например, ссылаясь на текстовое окно 29 для возможных сочетаний слов на Фиг.1, некоторые из перечисленных возможных сочетаний слов могут быть получены в результате поиска с универсальным шаблоном, где символ универсального шаблона предшествует группе "this person" (этого человека). Поисковым термином может быть, например, выражение "* this

5 person", где звездочка "*" представляет универсальный шаблон, и результаты включают в себя выражения "known this person", "admitted this person", "thought this person" и т.д. Выбор символа, используемого для универсального шаблона, произволен, и символом универсального шаблона может быть амперсанд, знак вопроса или какой-либо другой знак.

10 Универсальный шаблон может использоваться в каждом из ряда терминов опроса, основанных на одном и том же сочетании слов, где поиски выполняются от более ограниченных шаблонов запроса, основанных на этом сочетании слов, к менее ограниченным шаблонам запроса, аналогично тому, чтобы было сделано в первую очередь для оценки правильности или нестандартности заданного сочетания слов. Например, для предложения с неправильным сочетанием слов "I have recognized this person for years" ряд терминов запроса с универсальным шаблоном может включать в себя выражение "I have * this person for years"; один или более терминов запроса, 15 основанных на формате сокращенного предложения, то есть "I have * this person", "have * this person for years", "I have * this person"; один или более терминов запроса, основанных на паре отрывков, содержащих заданное сочетание слов, то есть "* this person for years", "* this person"; и один или более терминов запроса, основанных на паре слов, поиск которых выполняется либо для условия, когда они расположены рядом друг с другом, либо для условия, когда они расположены вблизи друг друга и отделены друг от друга не более чем одним промежуточным словом, например, "* ~ person", где знак тильды "~" в одном варианте осуществления произвольно выбирается для интерпретации поисковым механизмом в качестве заполнителя интервала, 20 25 который может заполнять интервал с длиной от нуля до одного слова или нескольких слов, отделяющих универсальный шаблон от специфицированного слова.

Соответственно, применение универсального шаблона в термине запроса предоставляет один иллюстративный пример эффективного способа сравнения 30 сочетаний слов из образца текста с потенциально аналогичными сочетаниями слов на различных уровнях в содержимом, доступном в сети Интернет или другом сетевом или локальном корпусе.

Также может использоваться специализированный универсальный шаблон, 35 который ограничен конечным набором предпочтительных возможных вариантов, либо эталон сочетаний слов, с которым сравниваются результаты, может включать в себя индикацию предпочтительных возможных сочетаний слов. Например, определенным возможным сочетаниям слов может быть отдано предпочтение на 40 45 основании процесса обратного перевода с родного языка автора заданного образца текста. Пользователю может быть предоставлена возможность ввода родного языка, или программное обеспечение проверки сочетаний слов может использовать инструмент для оценки признаков в образце, которые могут указывать вероятный родной язык автора, например, и пользователю может быть выведено приглашение для подтверждения выбора языка. Также могут рассматриваться другие признаки, 50 такие как языковые настройки, которые используются в другом программном обеспечении на компьютере, или индикации о физическом расположении компьютера, например. Например, если программное обеспечение детектирует, что оно выполняется на компьютере, на котором действует китайская версия операционной системы, или если программное обеспечение детектирует, что компьютер присоединен к сети Интернет через сервер, который расположен в Пекине, и если образец текста составлен на английском языке, то программное обеспечение может назначить предпочтительную или приоритетную обработку возможных сочетаний слов, которые

типовы для носителя китайского языка.

Если известен родной язык автора, то программное обеспечение проверки сочетаний слов может проверить, основаны ли неправильные сочетания слов на некорректном сопоставлении родного языка автора с неродным языком, на котором написан образец текста. Это может включать в себя, например, ошибки сопоставления при применении грамматического правила родного языка в неродном языке, неправильное применение слов, схожих по произношению или написанию, но имеющих другое значение, пропуск множественной формы, когда в родном языке автора не различают множественной/единственной формы и т.п.

Например, может быть определено, что автор фразы "I have recognized this person for years" является носителем китайского языка. В результате программное обеспечение проверки сочетаний слов может перевести этот отрывок на китайский, и проверить возможные переводы обратно на английский. Слово "To recognize" может быть переведено на китайский как "rènshí" (транслитерация согласно системе Пиньинь) - слово, которое может быть переведено на английский либо как "to recognize" (узнавать), либо как "to know" (знать). Следовательно, обратный перевод может указать, что слово "know" (знать) является возможной заменой слова "recognize" (узнавать), и перевод принимает следующий вид "I have known this person for years" (Я знаю этого человека много лет). В качестве еще одного примера образец текста на английском, автор которого по некоторым признакам является носителем французского, может содержать предложение "I commanded the faith and onions" (Я приказал веру и лук", где словам "commanded" (приказал) и "faith" (вера) присвоены теги нестандартных сочетаний слов. Это может быть переведено на французский как "J'ai commandé le foi et des oignons", где слово "commandé" правильно переводится обратно как слово "ordered" (заказал), а слово "le foi" пишется как "le foie", что переводится обратно как "the liver" (печень), так что программное обеспечение расставляет по приоритету или присваивает весовые коэффициенты этим возможным сочетаниям слов из всех других сочетаний, полученных в результате поиска, и могут быть предложены возможные заменяющие сочетания слов, в результате чего формируется предложение "I ordered the liver and onions" (Я заказал печень и лук).

В любом случае обратный перевод с родного языка может использоваться для ограничения поиска возможной замены в первую очередь или он может быть использован в качестве эталона сочетания слов для оценки или расстановки по приоритету потенциальных возможных замен, генерируемых в результате поиска.

После выполнения поиска эталона сочетаний слов для терминов запроса, как на этапе 73, в этом иллюстративном варианте осуществления может быть выполнен этап 75 для идентификации результатов поиска с относительно большой долей, в которых возможное слово заменяет универсальный шаблон. В одном иллюстративном варианте осуществления также используется присвоение весовых коэффициентов для различных типов совпадений, включающих в себя совпадения целого предложения, сокращенного предложения, пары отрывков или пары слов, которые включают в себя заданное сочетание слов. Общая оценка может представлять собой взвешенную сумму оценочных компонентов по каждому типу терминов запроса, где каждый оценочный компонент, например, представляет собой произведение количества результатов поиска, найденных для данного термина запроса, и весового коэффициента, присвоенного данному термину запроса. В этом иллюстративном варианте осуществления относительно высокая общая оценка, полученная согласно подобному способу, может быть интерпретирована как

эффективная мера относительно большой доли сочетаний, в которых возможное слово заменяет универсальный шаблон, и служит как вероятная возможность для коррекции неправильного сочетания.

Универсальный шаблон может быть конкретно нацелен на определенные части речи или типы отрывков в термине запроса. Например, в наборе терминов запроса, основанных на сочетаниях слов глагол-существительное, глаголы в каждом из сочетаний слов могут быть заменены на универсальный шаблон в одном наборе терминов запроса, и существительные в каждом из сочетаний слов могут быть заменены на универсальный шаблон в другом наборе терминов запроса. Для сочетаний слов предлог-существительное предлоги, в частности, могут представлять собой часть сочетания, выбранную для замены на универсальный шаблон, поскольку существительное более вероятно будет связано со значением предложения, и конкретный предлог, который должен быть выбран, может быть использован изучающими иностранный язык в неправильном сочетании слов. Следовательно, выбор предлога для замены на универсальный шаблон в терминах запроса, основанных на сочетаниях слов предлог-существительное, может поддерживать стратегию для сокращения времени, необходимого для образования желаемого замещающего сочетания слов. Аналогично, как универсальный шаблон может быть предпочтительным образом выбрано прилагательное в сочетании слов прилагательное-существительное, или наречие в сочетании слов глагол-наречие, поскольку существительное и глагол наиболее вероятно будут связаны с намеченным значением, тогда как прилагательные и наречия могут быть подвержены возможности неправильного использования в сочетании слов в большей степени, чем существительные и глаголы, которые они модифицируют. В других вариантах осуществления для предпочтительной или исключительной замены на универсальные шаблоны могут быть выбраны другие части речи или типы отрывков.

Так, потенциальное замещающее слово, которое имеет значительно большую долю, чем подозрительное слово, может быть указано как предполагаемая замена для нестандартной части сочетания слов. Это может быть сделано после выполнения перекрестной ссылки сочетания слов и потенциального замещающего слова для включения в эталон сочетаний слов, такой как словарь правильных сочетаний слов, или перекрестной ссылки с предлагаемыми обратными переводами, основанными на родном языке автора, как описано выше.

Результаты поиска, имеющие одно или несколько потенциально возможных замещающих слов, таких как возможные слова с наивысшей оценкой совпадения, могут быть предоставлены пользователю через устройство вывода, как потенциально правильные сочетания слов, как на этапе 77 на Фиг.2. Это может быть сделано аналогично возможным сочетаниям слов, отображенными в текстовом окне 29 на Фиг.1, например, так что пользователю предоставляется возможность выбрать одно из возможных сочетаний слов, чтобы заменить сочетание, которое обозначено как неправильное. Иногда, для одного из сочетаний слов вероятность правильной замены неправильного сочетания слов будет существенно выше. Этот случай показан на Фиг.4.

Фиг.4 представляет собой иллюстрацию пользовательского интерфейса 410 для вычислительной системы, реализующей способ проверки согласно иллюстративному варианту осуществления настоящего изобретения. В верхней части пользовательского интерфейса 410 присутствуют те же вкладки, что и у интерфейса 10 проверки ошибок сочетаний слов на Фиг.1, то есть вкладка 11 "Загрузить Предложения ESL" и

вкладка 13 "Выполнить Запросы". Фиг.1 иллюстрирует пользовательский интерфейс 10 при выбранной вкладке 11 "Загрузить Предложения ESL", а Фиг.4 иллюстрирует пользовательский интерфейс 410 при выбранной вкладке 13 с соответствующими кнопками, заголовками, текстовыми окнами и другими 5 элементами пользовательского интерфейса. Пользовательский интерфейс 410 разделен на верхнюю секцию "Проанализированные Предложения ESL" и нижнюю секцию "Уровни Запросов". Пользовательский интерфейс 410 предназначен для поиска по конкретным типам запросов и для предоставления через пользовательский интерфейс 10 результатов поиска, в поддержку иллюстративного варианта осуществления проверки ошибок сочетаний слов.

Верхняя секция "Проанализированные Предложения ESL" включает в себя текстовое окно 25, которое аналогично текстовому окну 25 на Фиг.1 и которое в данном примере содержит то же проанализированное предложение. Эта секция также 15 включает в себя ряд 41 кнопок под текстовым окном 25 проанализированного предложения. Эти кнопки предназначены для выполнения поиска по различным иллюстративным шаблонам запроса, основанным на данном проанализированном предложении. В частности, в этом иллюстративном варианте осуществления 20 присутствуют кнопки "Создать Запросы"; кнопки "Передать П-Запросы I" и "Передать П-Запросы II", которые относятся к различным запросам на основе предложения, как, например, для целого или сокращенного предложения; кнопки "Передать К-Запросы I" и "Передать К-Запросы II" для различных типов запросов по паре отрывков и кнопка "Передать С-Запросы" для запросов, основанных на паре 25 отдельных слов заданного сочетания, в соответствии с описанным выше.

Нижняя секция пользовательского интерфейса 410 включает в себя текстовое окно 43 для запросов уровня предложения, текстовое окно 45 для запросов первого уровня отрывка, и текстовое окно 47 для запросов второго уровня отрывка. 30 Упомянутые окна представляют иллюстративные примеры текстовых окон для результатов запросов, которые могут быть вызваны для представления возможных правильных слов для замены. В каждом из текстовых окон 43, 45 и 47 также показан счетчик совпадений результатов поиска по запросу для этого возможного сочетания, а с правой стороны упомянутых текстовых окон 43, 45 и 47 отображены отдельные 35 текстовые окна, в которых могут выводиться некоторые результаты поиска из корпуса, соответствующие каждому типу запроса. На этой фигуре можно заметить, что слово "known" (знал) строго указывается как правильная замена слова "recognized" (узнал) на каждом уровне шаблона запроса. Это иллюстрирует, что многоуровневая 40 схема шаблонов запроса предоставляет избыточность, которой в этом случае будет более чем достаточно для представления убедительного случая для одной конкретной возможной замены. Тогда пользователь может учесть количество совпадений для каждого шаблона запроса и выбрать одно из возможных заменяющих сочетаний, чтобы заменить нестандартное сочетание слов, которому оно соответствует.

45 Наряду с тем, что Фиг.1 и 4 иллюстрируют панели графического отображения, связанные с вариантом осуществления специализированного приложения, другой вариант осуществления может функционировать менее очевидным образом в привязке к другому приложению, такому как программа обработки текста, веб-браузер, приложение электронной почты, программа для презентаций или другое приложение, 50 в которое пользователь вводит текст или другую форму естественного языка. Подобная программа или приложение может взаимодействовать с отдельной программой проверки ошибок сочетаний слов или может включать в себя

собственный модуль, который содержит вариант осуществления проверки ошибок сочетаний слов, например. В других вариантах осуществления, например, инструмент проверки ошибок сочетаний слов может быть доступен через меню "Сервис" в другом приложении или он может быть доступен как опция во всплывающем меню, когда 5 пользователь выполняет правый щелчок на слове.

В еще одном варианте осуществления некая другая программа может по умолчанию выполнять модуль или программу проверки ошибок сочетаний слов автоматически и может автоматически предоставлять индикации неправильных или 10 нестандартных сочетаний слов посредством маркирования слов на экране. Например, упомянутое маркирование может иметь форму синей зигзагообразной линии под сочетанием слов, которое определяется как неправильное или нестандартное. В еще одном варианте осуществления может выполняться отдельный модуль или 15 программа, чтобы детектировать, присутствуют ли во вводе пользователя признаки того, что пользователь не является носителем языка ввода, и эти индикации, например, могут инициировать активацию модуля или программы проверки ошибок сочетаний слов.

Фиг.5 представляет собой пример подходящей среды 100 вычислительной системы, 20 в котором могут быть реализованы различные варианты осуществления настоящего изобретения. Например, различные варианты осуществления могут быть реализованы как программные приложения, модули или другие формы инструкций, которые выполняются средой 100 вычислительной системы и которые конфигурируют среду 100 вычислительной системы таким образом, чтобы выполнять различные 25 задачи или способы различных вариантов осуществления. Программное приложение или модуль, реализующий вариант осуществления проверки ошибок сочетаний слов, может быть разработан на любом из языков или окружений программирования или сценариев. Например, он может быть написан на языке C#, F#, C++, C, Pascal, Visual 30 Basic, Java, JavaScript, Delphi, Eiffel, Nemerle, Perl, PHP, Python, Ruby, Visual FoxPro, Lua или любом другом языке программирования. Также предполагается, что языки программирования и другие формы создания выполняемых инструкций продолжат развиваться, и с их помощью могут быть разработаны дополнительные варианты осуществления.

Согласно одному иллюстративному варианту осуществления среда 100 35 вычислительной системы может быть сконфигурирована таким образом, чтобы выполнять задачи проверки ошибок сочетаний слов в ответ на получение индикации сочетания слов в тексте. Тогда окружение 100 вычислительной системы может выполнить в сети Интернет поиск одного или более шаблонов запроса, связанных с 40 указанным сочетанием слов. Различные используемые шаблоны запросов могут включать в себя предложение, сокращенное предложение, пару отрывков и/или пару отдельных слов, любое из которых может включать в себя сочетание слов. Затем среда 100 вычислительной системы может оценить результаты поиска в сети Интернет 45 для каждого из шаблонов запроса и определить, соответствует ли заданное сочетание слов нормальному использованию или оно является нестандартным или указывает о вероятной ошибке. Нормальное использование может быть указано либо наличием точного совпадения шаблона запроса, содержащего заданное предложение, либо 50 оценкой совпадения, которая больше предопределенного порогового значения. Далее, система может указать, как часть вывода среды 100 вычислительной системы через устройство вывода в виде результата работы варианта осуществления способа проверки ошибок сочетания слов, соответствует ли сочетание словциальному

использованию или оно является нестандартным и указывается как неправильное.

Показанная на Фиг.5 среда 100 вычислительной системы является лишь одним примером подходящего вычислительного окружения для выполнения и предоставления вывода из различных вариантов осуществления, и оно не предназначено для определения рамок объема использования или функционирования сущности изобретения, определенного формулой изобретения. Кроме того, вычислительная среда 100 не должна интерпретироваться как имеющая зависимость или требования, относящиеся к какому-либо компоненту или комбинациям компонентов, проиллюстрированных в примере рабочей среды 100.

Варианты осуществления могут быть реализованы с множеством других сред или конфигураций сред вычислительной системы общего назначения или специального назначения. Примеры известных вычислительных систем, сред и/или конфигураций, которые могут подходить для использования с различными вариантами осуществления, включают в себя, но не ограничиваются перечисленным, персональные компьютеры, серверные компьютеры, карманные или портативные устройства, многопроцессорные системы, системы, основанные на микропроцессорах, сетевые персональные компьютеры, миникомпьютеры, универсальные компьютеры (мэйнфреймы), системы телефонии, распределенные вычислительные окружения, которые включают в себя любую из вышеперечисленных систем или устройств, и т.п.

Варианты осуществления могут быть описаны в общем контексте выполняемых компьютером инструкций, таких как программные модули, которые выполняются компьютером. В общем, программные модули включают в себя рутинные процедуры, программы, объекты, компоненты, структуры данных и т.п., которые выполняют конкретные задачи или осуществляют конкретные абстрактные типы данных.

Некоторые варианты осуществления могут быть реализованы так, чтобы использоваться в распределенных вычислительных средах, где задачи выполняются посредством удаленных устройств обработки, которые связаны через сеть связи. В распределенной вычислительной среде программные модули расположены как на носителе данных локального компьютера, так и на носителе данных удаленного компьютера, включающем в себя устройства памяти. Как описано в настоящем документе, такие выполняемые инструкции могут храниться на носителе данных, так что они могут быть считаны и выполнены одним или более компонентами вычислительной системы, таким образом, предоставляя вычислительной системе новые возможности.

Со ссылкой на Фиг.5, пример системы для осуществления некоторых вариантов осуществления настоящего изобретения включает в себя вычислительное устройство общего назначения в форме компьютера 110. Компоненты компьютера 110 могут включать в себя, но не ограничены перечисленным, процессорный блок 120, системную память 130 и системную шину 121, которая соединяет различные компоненты системы, в том числе системную память с процессорным блоком 120.

Системная шина 121 может быть любого типа из ряда типов структур шин, включающих в себя шину памяти или контроллер памяти, периферийную шину и локальную шину, используя любую архитектуру из разнообразия архитектур шин. В качестве примера и не ограничиваясь перечисленным, подобные архитектуры включают в себя шину стандарта Industry Standard Architecture (ISA), шину стандарта Micro Channel Architecture (MCA), шину стандарта Enhanced ISA (EISA), локальную шину стандарта Video Electronics Standards Association (VESA) и шину стандарта Peripheral Component Interconnect (PCI), также известную как шина

расширения.

Компьютер 110, как правило, включает в себя ряд машиночитаемых носителей. Машиночитаемые носители могут представлять собой любые доступные носители, к которым компьютер 110 может выполнить доступ, и они включают в себя как 5 энергозависимые, так и энергонезависимые носители, съемные и несъемные носители. В качестве примера, но не ограничиваясь перечисленным, машиночитаемые носители могут содержать компьютерные носители хранения и среды связи. Компьютерные 10 носители хранения включают в себя энергозависимые, энергонезависимые, съемные и несъемные носители, реализованные посредством какого-либо способа или 15 технологии для хранения информации, такой как машиночитаемые команды, структуры данных, программные модули и другие данные. Компьютерные носители хранения включают в себя, но не ограничены этим, ОЗУ, ПЗУ, ЭСППЗУ, флэш-память или другую технологию памяти, диски CD-ROM, цифровые универсальные 20 диски (DVD) или иные оптические дисковые носители, магнитные кассеты, магнитные ленты, магнитные дисковые носители или другие магнитные устройства хранения или любой другой носитель, который может быть использован, чтобы хранить желаемую 25 информацию, и к которой может быть выполнен доступ компьютером 110. Среды связи, как правило, заключают в себя машиночитаемые команды, структуры данных, 30 программные модули и другие данные в виде модулированного сигнала данных, такого как несущая волна или другой механизм передачи, и включают в себя любые среды доставки информации. Термин "модулированный сигнал данных" обозначает сигнал, у которого одна или более характеристик установлены или изменяются таким образом, чтобы кодировать в сигнал информацию. В качестве примера, но не ограничиваясь перечисленным, среды связи включают в себя проводные среды, такие как проводная сеть или прямое проводное соединение, и беспроводные среды, такие как акустические, радиочастотные, инфракрасные и другие беспроводные среды.

Комбинации из каких-либо вышеперечисленных типов также входят в объем понятия машиночитаемый носитель.

Системная память 130 включает в себя компьютерные носители хранения в форме энергозависимой и/или энергонезависимой памяти, такой как ПЗУ 131 и ОЗУ 132.

Базовая система 133 ввода/вывода (BIOS), содержащая базовые рутинные процедуры, 35 которые помогают передавать информацию между элементами в компьютере 110, как например, во время загрузки, как правило, хранится в ПЗУ 131. ОЗУ 132, как правило, содержит данные и/или программные модули, которые непосредственно доступны и/или задействованы процессорным блоком 120. В качестве примера, но не ограничиваясь этим, Фиг.5 иллюстрирует операционную систему 134, прикладные 40 программы 135, другие программные модули 136 и программные данные 137.

Компьютер 110 может также включать в себя другие съемные/несъемные энергозависимые/энергонезависимые компьютерные носители информации.

Исключительно в качестве примера, Фиг.5 иллюстрирует привод 141 жесткого диска, 45 который считывает с или записывает на несъемный, энергонезависимый магнитный носитель, привод 151 магнитного диска, который считывает с или записывает на съемный, энергонезависимый магнитный диск 152, и привод 155 оптического диска, который считывает с или записывает на съемный, энергонезависимый оптический 50 диск 156, такой как CD-ROM или другой оптический носитель информации. Другие съемные/несъемные, энергозависимые/энергонезависимые компьютерные носители информации, которые могут быть использованы в примере рабочего окружения, включают в себя, но не ограничиваются перечисленным, кассеты с магнитной лентой,

5 карты флэш-памяти, цифровые универсальные диски, цифровые видеоленты, твердотельные ОЗУ, твердотельные ПЗУ и т.п. Привод 141 жесткого диска, как правило, соединен с системной шиной 121 через интерфейс несъемной памяти, такой как интерфейс 140, а привод 151 магнитного диска и привод 155 оптического диска, как правило, соединены с системной шиной 121 посредством интерфейса съемной памяти, такого как интерфейс 150.

10 Приводы и связанные с ними компьютерные носители информации, описанные выше и проиллюстрированные на Фиг.5, предоставляют хранение машиночитаемых команд, структур данных, программных модулей и других данных для компьютера 110. На Фиг.5, например, привод 141 жесткого диска проиллюстрирован, как хранящий операционную систему 144, прикладные программы 145, другие 15 программные модули 146 и программные данные 147. Следует отметить, что эти компоненты могут быть такими же, как операционная система 134, прикладные программы 135, другие программные модули 136 и программные данные 137 или же отличаться от них. Операционная система 144, прикладные программы 145, другие 20 программные модули 146 и программные данные 147 обозначены различными номерами, чтобы проиллюстрировать, что, по меньшей мере, они представляют собой различные копии.

25 Пользователь может вводить команды и информацию в компьютер 110 посредством устройств ввода, таких как клавиатура 162, микрофон 163 и указывающее устройство 161, такое как мышь, трекбол или сенсорная панель. Другие 30 устройства ввода (не показаны) могут включать в себя джойстик, игровой планшет, спутниковую антенну, сканер или т.п. Эти и другие устройства ввода часто соединяются с процессорным блоком 120 через интерфейс 160 ввода пользователя, который соединен с системной шиной, но они могут также быть соединены 35 посредством другого интерфейса и структур шины, такой как параллельный порт, игровой порт или универсальная последовательная шина (USB). Монитор 191 или другой тип устройства отображения также соединен с системой шиной 121 посредством интерфейса, такого как видеоинтерфейс 190. В добавление к монитору, 40 компьютеры могут также включать в себя другие периферийные устройства вывода, такие как громкоговорители 197 и принтер 196, которые могут быть соединены через интерфейс 195 периферийных устройств вывода.

Компьютер 110 работает в сетевом окружении, используя логические соединения с одним или более удаленными компьютерами, такими как удаленный компьютер 180. Удаленный компьютер 180 может быть персональным компьютером, карманным 45 устройством, сервером, маршрутизатором, сетевым персональным компьютером, устройством однорангового узла или другим обычным сетевым узлом, и он, как правило, включает в себя многие или все элементы, описанные выше относительно компьютера 110. Логические соединения, изображенные на Фиг.5, включают в себя локальную сеть (Local Area Network, LAN) 171 и глобальную сеть (Wide Area Network, 50 WAN) 173, но могут также включать в себя другие сети. Подобные сетевые окружения типичны для учреждений, компьютерных сетей масштаба предприятия, интранета и Интернета.

При использовании в среде локальной сети компьютер 110 соединен с локальной 55 сетью 171 через сетевой интерфейс или адаптер 170. При использовании в среде глобальной сети компьютер 110, как правило, включает в себя модем 172 или иное средство для установления связи через глобальную сеть 173, такую как Интернет. Модем 172, который может быть внутренним или внешним, может быть соединен с

системной шиной 121 посредством интерфейса 160 ввода пользователя или иного подходящего механизма. В сетевой среде программные модули, изображенные относительно компьютера 110, или их части могут храниться в удаленном устройстве памяти. В качестве примера, но не ограничиваясь этим, Фиг.5 иллюстрирует удаленные прикладные программы 185 как находящиеся на удаленном компьютере 180. Очевидно, что показанные сетевые соединения представляют собой лишь примеры и могут быть использованы другие средства для установления линии связи между компьютерами.

На Фиг.6 проиллюстрирована структурная схема мобильной вычислительной среды, содержащей мобильное вычислительное устройство и носитель, который может быть считан мобильным вычислительным устройством и который содержит выполняемые инструкции, которые могут быть выполнены мобильным вычислительным устройством согласно еще одному иллюстративному варианту осуществления. На Фиг.6 показана структурная схема мобильной вычислительной системы 200, включающей в себя мобильное устройство 201 согласно иллюстративному варианту осуществления. Мобильное устройство 201 включает в себя микропроцессор 202, память 204, компоненты 206 ввода/вывода и интерфейс 208 связи для осуществления связи с удаленными компьютерами или другими мобильными устройствами. В одном варианте осуществления вышеупомянутые компоненты соединены друг с другом через подходящую шину 210.

Память 204 реализована как энергонезависимая электронная память, такая как ОЗУ с резервным аккумуляторным модулем (не показан), так что хранимая в памяти 204 информация не утрачивается, когда общее питание мобильного устройства 200 отключается. В качестве примера часть памяти 204 выделена как адресуемая память для выполнения программ, тогда как другая часть памяти 204 используется для хранения, как например, для имитации хранилища на жестком диске.

Память 204 включает в себя операционную систему 212, прикладные программы 214, а также хранилище 216 объектов. Во время работы операционная система 212 считывается из памяти 204 и выполняется процессором 202. В одном иллюстративном варианте осуществления операционная система 212 представляет собой операционную систему WINDOWS® CE корпорации Microsoft. Операционная система 202 спроектирована для мобильных устройств, и она реализует функции базы данных, которая может быть использована приложениями 214 через ряд интерфейсов и способов прикладного программирования. Объекты в хранилище 216 объектов поддерживаются приложениями 214 и операционной системой 212, по меньшей мере, частично, в ответ на вызовы программируемых интерфейсов и способов подвергающегося воздействию приложения.

Интерфейс 208 связи представляет множество устройств и технологий, которые позволяют мобильному устройству 200 передавать и принимать информацию. Эти устройства включают в себя проводные и беспроводные модемы, спутниковые приемники, приемники широкополосной передачи и т.п. Мобильное устройство 200 также может быть напрямую соединено с компьютером, чтобы обмениваться с ним данными. В подобных случаях интерфейс 208 связи может представлять собой инфракрасный приемопередатчик или последовательное или параллельное соединение, которые способны передавать поток информации.

Компоненты 206 ввода/вывода включают в себя множество разных устройств ввода, таких как сенсорный экран, кнопки, ползунки и микрофон, а также различные устройства вывода, включающие в себя генератор звуков, вибрирующее устройство и

дисплей. Вышеперечисленные устройства представляют собой только пример, и наличие всех этих устройств в мобильном устройстве 200 необязательно. В добавление, к мобильному устройству 200 могут быть подключены другие устройства ввода/вывода.

5 Мобильная вычислительная система 200 также включает в себя сеть 220. В данном примере мобильное вычислительное устройство 201 находится в беспроводной связи с сетью 220 (которая может представлять собой Интернет, глобальную сеть или локальную сеть, например) путем передачи и приема электромагнитных сигналов 299

10 подходящего протокола между интерфейсом 208 связи и беспроводным интерфейсом 222. Беспроводной интерфейс 222 может представлять собой, например, беспроводной концентратор или сотовую антенну или любой другой сигнальный интерфейс. В свою очередь беспроводной интерфейс 222 предоставляет доступ через сеть 220 к множеству дополнительных вычислительных ресурсов, иллюстративно

15 представленных вычислительными ресурсами 224 и 226. Само собой разумеется, что с сетью 220 может осуществлять связь любое количество вычислительных устройств. Вычислительному устройству 201 предоставляется возможность использовать выполняемые инструкции, хранимые на носителе запоминающего компонента 204,

20 такие как выполняемые инструкции, которые в нескольких иллюстративных вариантах осуществления позволяют вычислительному устройству 201 выполнять запросы проверки неродного языка и другие задачи.

Несмотря на то, что сущность настоящего изобретения была описана в привязке к структурным особенностям и/или методологическим действиям, следует понимать, что 25 сущность, определенная в прилагаемой формуле изобретения, не ограничена конкретными особенностями или действиями, описанными выше. Скорее, описанные выше конкретные структурные особенности и действия раскрыты как примеры форм осуществления пунктов формулы изобретения. В частности, несмотря на то, что в 30 настоящем документе для удобства используются термины "компьютер", "вычислительное устройство" или "вычислительная система", следует понимать, что каждый из данных терминов может обозначать любое вычислительное устройство, вычислительную систему, вычислительную среду, мобильное устройство или другой компонент или контекст обработки информации, и они не ограничиваются какой-либо 35 отдельной интерпретацией. В качестве еще одного конкретного примера, наряду с тем, что многие варианты осуществления представлены посредством иллюстративных элементов, которые широко известны на момент написания настоящей патентной заявки, предполагается, что множество новых инноваций в компьютерной технологии 40 повлияют на различные варианты осуществления в таких аспектах, как пользовательские интерфейсы, способы пользовательских вводов, вычислительные среды и вычислительные способы, и что элементы, определенные пунктами формулы изобретения, могут быть реализованы согласно этим и другим инновациям, между тем оставаясь в рамках определений формулы изобретения.

45

Формула изобретения

1. Способ проверки ошибок сочетаний слов, реализуемый вычислительной системой, содержащей один или более процессоров, содержащий этапы, на которых сравнивают, с использованием одного или более процессоров, одно или более сочетаний слов из образца текста с корпусом;

50 идентифицируют, с использованием одного или более процессоров, являются ли упомянутые сочетания слов нестандартными в упомянутом корпусе; и

через устройство вывода предоставляют индикации о том, являются ли упомянутые сочетания слов нестандартными;

при этом на этапе сравнения упомянутых сочетаний слов с упомянутым корпусом выполняют один или более поисков в Интернете, используя один или более терминов запроса, которые содержат каждое из одного или более упомянутых сочетаний слов;

при этом для каждого из одного или более упомянутых сочетаний слов, для которых выполняются поиски, выполняется поиск для каждого из одного или более терминов запроса, которые содержат упомянутые сочетания слов, до тех пор, пока один из терминов запроса не предоставит результаты поиска, которые удовлетворяют предварительно выбранному пороговому значению для совпадения сочетания слов, или пока все термины запроса, которые содержат упомянутое сочетание слов, не будут использованы без достижения упомянутого предварительно выбранного порогового значения;

и дополнительно содержащий этапы, на которых составляют один или более терминов запроса с универсальным шаблоном, замещающим одно из слов в одном из нестандартных сочетаний слов;

выполняют поиск эталона сочетаний слов для упомянутых терминов запроса; идентифицируют результаты поиска с относительно большой долей замещения возможным словом универсального шаблона и

предоставляют через устройство вывода результаты поиска с возможным словом в качестве потенциально правильных сочетаний слов.

2. Способ по п.1, в котором корпус содержит содержимое в сети Интернет.

3. Способ по п.1, в котором сочетание слов является нестандартным, либо если оно отсутствует в корпусе, либо если его оценка соответствия меньше предварительно выбранного порогового значения, указывающего значительное присутствие нечетких совпадений упомянутого сочетания слов в упомянутом корпусе.

4. Способ по п.1, в котором корпус содержит содержимое, доступное в сети.

5. Способ по п.1, в котором для каждого из упомянутых сочетаний слов, для которых выполняются поиски, сочетание слов является нестандартным, если оценка терминов запроса, которые содержат упомянутое сочетание слов, не выше предварительно выбранного порогового значения оценки совпадения, указывающего значительное присутствие нечетких совпадений упомянутых терминов запроса в корпусе.

6. Способ по п.1, в котором по меньшей мере один из терминов запроса содержит предложение, которое содержит упомянутое сочетание слов.

7. Способ по п.1, в котором по меньшей мере один из терминов запроса содержит сокращенное предложение, которое содержит упомянутое сочетание слов, причем упомянутое сокращенное предложение формируется путем удаления вспомогательных слов из предложения, которое содержит упомянутое сочетание слов.

8. Способ по п.1, в котором по меньшей мере один из терминов запроса содержит пару отрывков, содержащих упомянутое сочетание слов.

9. Способ по п.1, в котором по меньшей мере один из терминов запроса содержит пару слов, содержащих упомянутое сочетание слов.

10. Способ по п.1, дополнительно содержащий этап, на котором маркируют частями речи образец текста и дополнительно в котором сочетания слов, которые сравниваются с упомянутым корпусом, выбирают из образца текста, содержащего по меньшей мере одно из сочетания слов глагол - существительное, сочетания слов предлог - существительное, сочетания слов прилагательное - существительное или

сочетания слов глагол - наречие.

11. Способ по п.1, в котором образец текста составлен на целевом языке и способ дополнительно содержит этап, на котором сканируют корпус, чтобы сравнивать заданные сочетания слов только с содержимым, составленным на целевом языке.

5 12. Способ по п.1, дополнительно содержащий этап, на котором сканируют образец текста, чтобы определить, присутствуют ли признаки использования неродного языка, и если в образце текста присутствуют признаки использования неродного языка, то автоматически инициируют этап, на котором сравнивают упомянутые сочетания слов

10 из образца текста с корпусом.

13. Способ по п.1, в котором термины запроса содержат один или более из шаблона предложения, шаблона отрывка и шаблона слова, причем результаты поиска с относительно большой долей замещения возможным словом универсального шаблона оцениваются путем перемножения доли результатов поиска, которые содержат

15 возможное слово, с весовым коэффициентом шаблона запроса, который имеет наибольшее значение для шаблона предложения, меньшее значение для шаблона отрывка и минимальное значение для шаблона слова.

14. Способ по п.1, в котором термины запроса содержат одно или более из

20 сочетания слов глагол - существительное, сочетания слов предлог - существительное, сочетания слов прилагательное - существительное и сочетания слов глагол - наречие и универсальный шаблон выбирается как глагол в сочетании слов глагол - существительное, как существительное в сочетании слов глагол - существительное, как предлог в сочетании слов предлог - существительное, как прилагательное в сочетании 25 слов прилагательное - существительное или как наречие в сочетании слов глагол - наречие.

15. Способ по п.1, дополнительно содержащий этап, на котором пользователю предоставляют возможность выбрать одно из потенциально правильных сочетаний

30 слов, чтобы заменить нестандартное сочетание слов, которому оно соответствует.

16. Читаемый компьютером носитель хранения, содержащий инструкции, выполняемые вычислительной системой, содержащей один или более процессоров, причем упомянутые инструкции конфигурируют упомянутую вычислительную систему таким образом, чтобы:

35 принимать индикацию сочетания слов в тексте;

выполнять, с использованием одного или более процессоров, поиск в сети Интернет для каждого из одного или более шаблонов запроса, связанных с указанным сочетанием слов, причем один из упомянутых шаблонов запроса содержит

40 предложение, в котором было обнаружено упомянутое сочетание слов, один из упомянутых шаблонов запроса содержит сокращенное предложение, основанное на предложении, в котором было обнаружено упомянутое сочетание слов, один из упомянутых шаблонов запроса содержит пару отрывков, содержащих упомянутое сочетание слов, и один из упомянутых шаблонов запроса содержит пару отдельных

45 слов, содержащих упомянутое сочетание слов; оценивать, с использованием одного или более процессоров, указывают ли результаты упомянутого поиска в сети Интернет для каждого из одного или более шаблонов запроса, что сочетание слов

50 соответствует нормальному использованию или является нестандартным сочетанием слов, что указывается либо точным совпадением шаблона запроса, содержащего предложение, либо точным совпадением шаблона запроса, содержащего сокращенное предложение, либо оценкой совпадения для шаблона запроса, содержащего пару отрывков, которая выше предварительно выбранного порогового значения для пары

отрывков, либо оценкой совпадения для шаблона запроса, содержащего пару отдельных слов, которая выше предварительно выбранного порогового значения для пары отдельных слов;

5 указывать, через воспринимаемое пользователем устройство вывода, соответствует ли упомянутое сочетание слов нормальному использованию или является ли упомянутое сочетание слов нестандартным;

и дополнительно для того, чтобы

составлять один или более терминов запроса с универсальным шаблоном,

10 замещающим слово в одном из нестандартных сочетаний слов;

выполнять поиск эталона сочетаний слов для упомянутых терминов запроса;

идентифицировать результаты поиска с относительно большой долей замещения возможным словом универсального шаблона и

15 предоставляемый через устройство вывода результаты поиска с возможным словом в качестве потенциально правильных сочетаний слов.

17. Вычислительная система, содержащая

компьютерный процессор и хранилище данных, доступное для компьютерного процессора, причем хранилище данных хранит компьютерно-читаемые инструкции и 20 компьютерный процессор получает доступ к компьютерно-читаемым инструкциям, чтобы:

идентифицировать сочетания слов в тексте;

выполнять в сети Интернет поиск набора шаблонов запроса, основанных на каждом из одного или более сочетаний слов; и

25 указывать через пользовательское устройство вывода, указывают ли результаты поиска, что упомянутые сочетания слов являются относительно редкими в сети Интернет;

составлять один или более терминов запроса с универсальным шаблоном,

30 замещающим слово в одном из упомянутых относительно редких сочетаний слов;

выполнять поиск в сети Интернет для упомянутых терминов запроса;

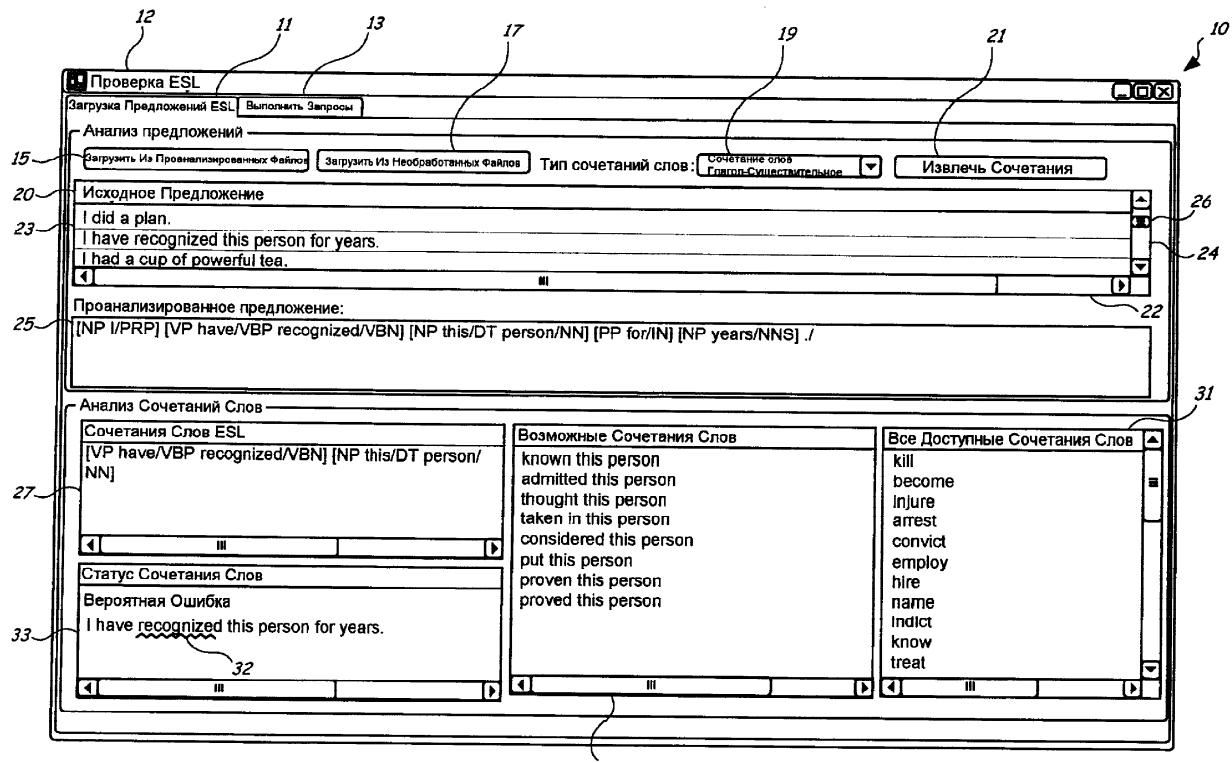
идентифицировать результаты поиска с относительно большой долей замещения возможным словом универсального шаблона и

35 предоставляемый через устройство вывода результаты поиска с возможным словом в качестве возможных замен для относительно редких сочетаний слов.

40

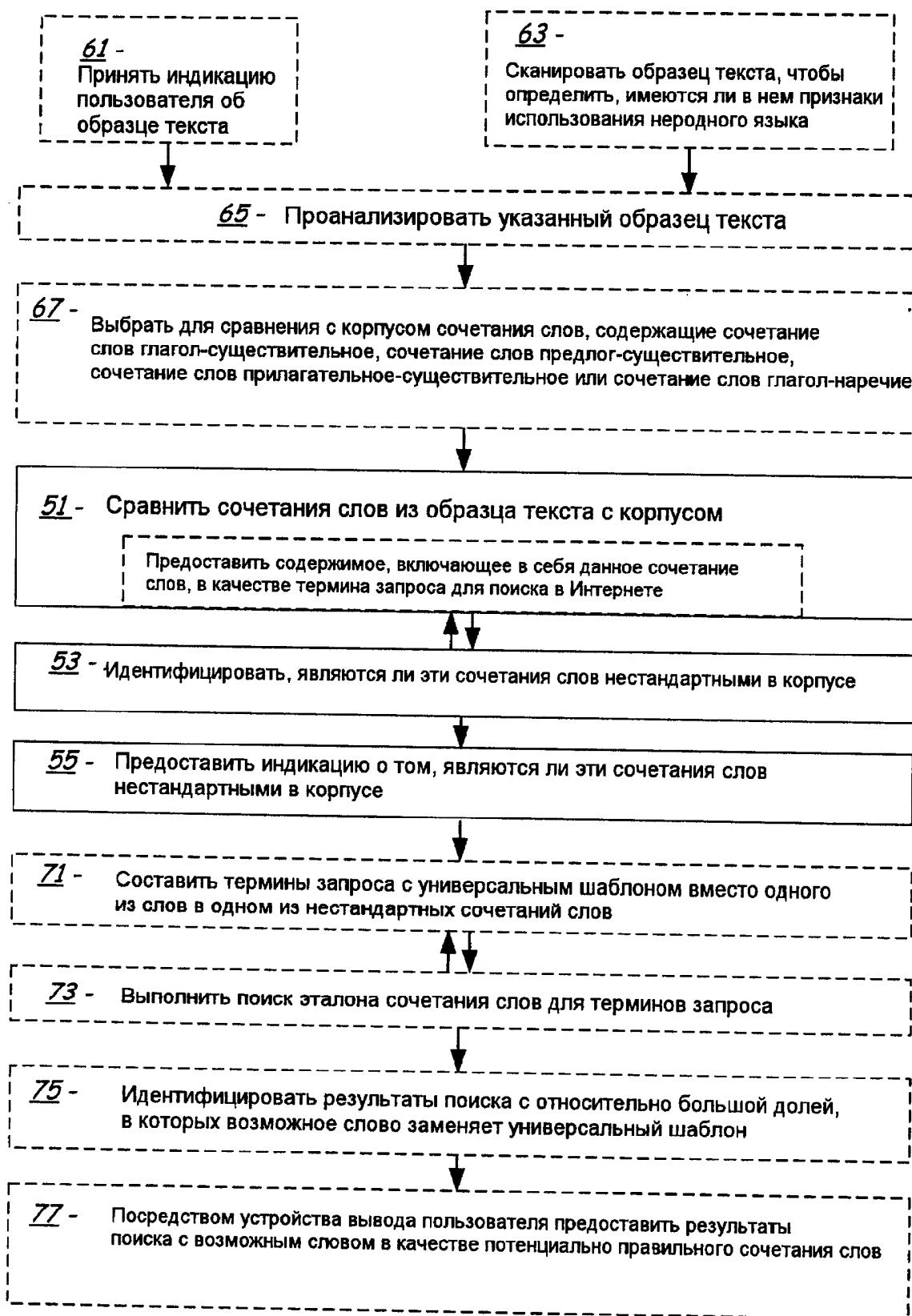
45

50

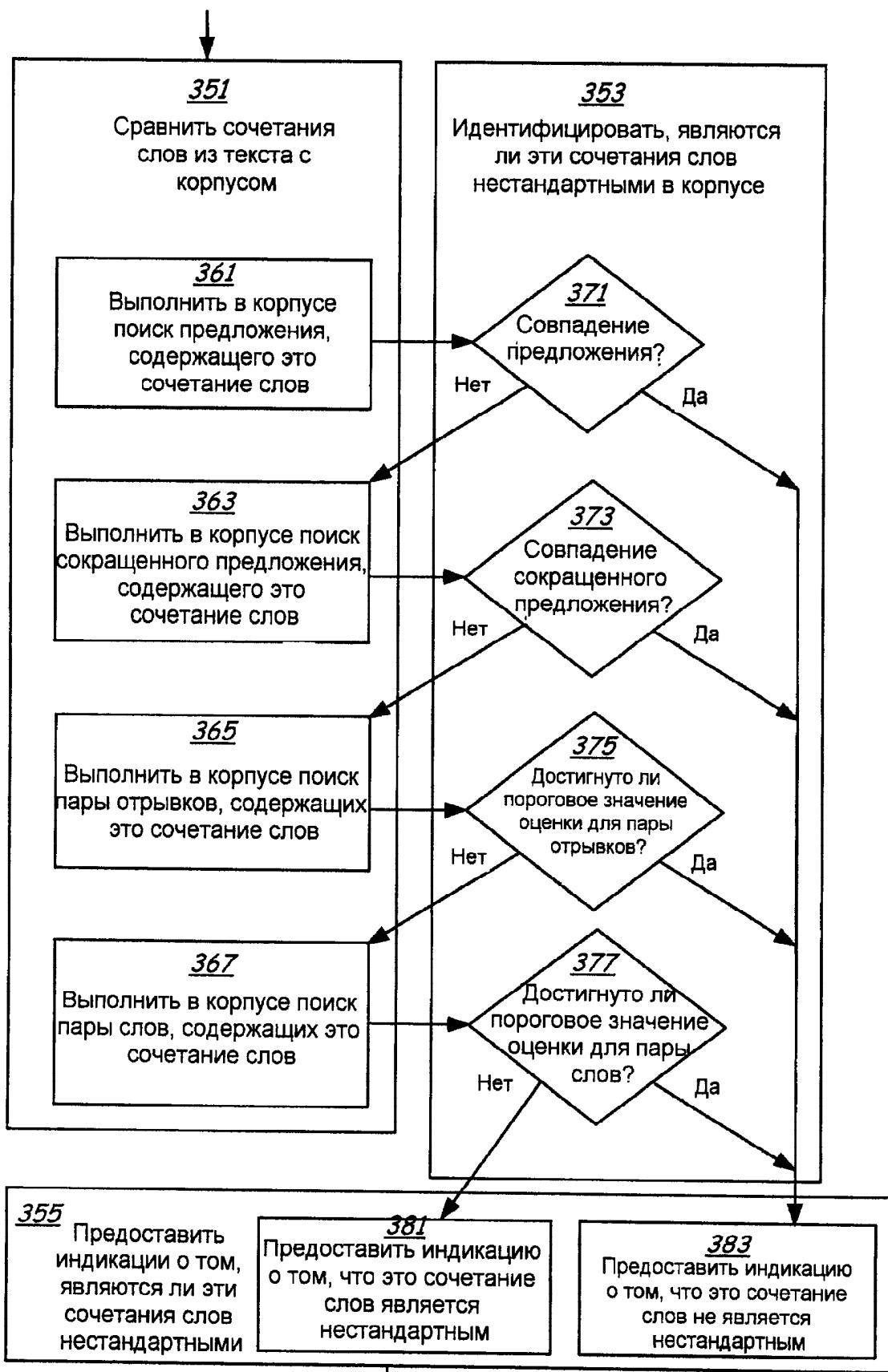


Фиг.1

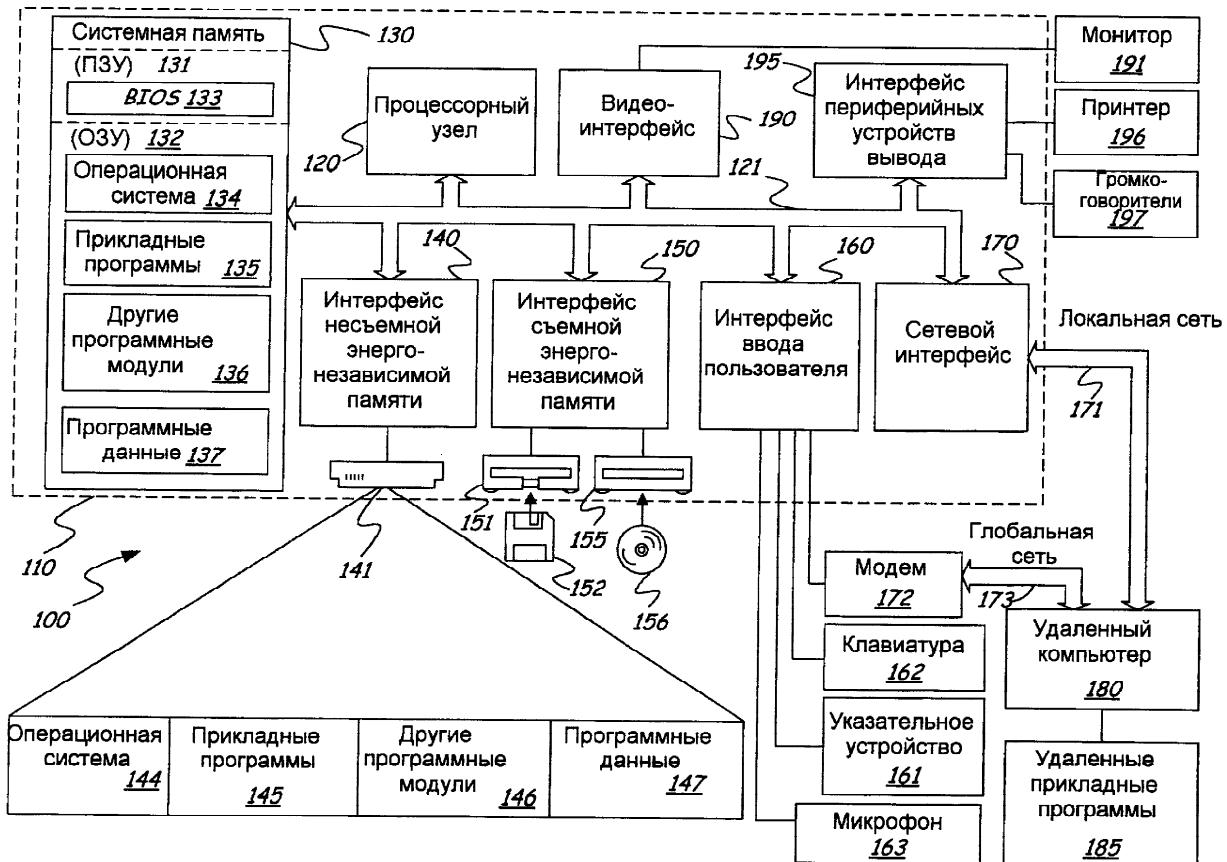
50



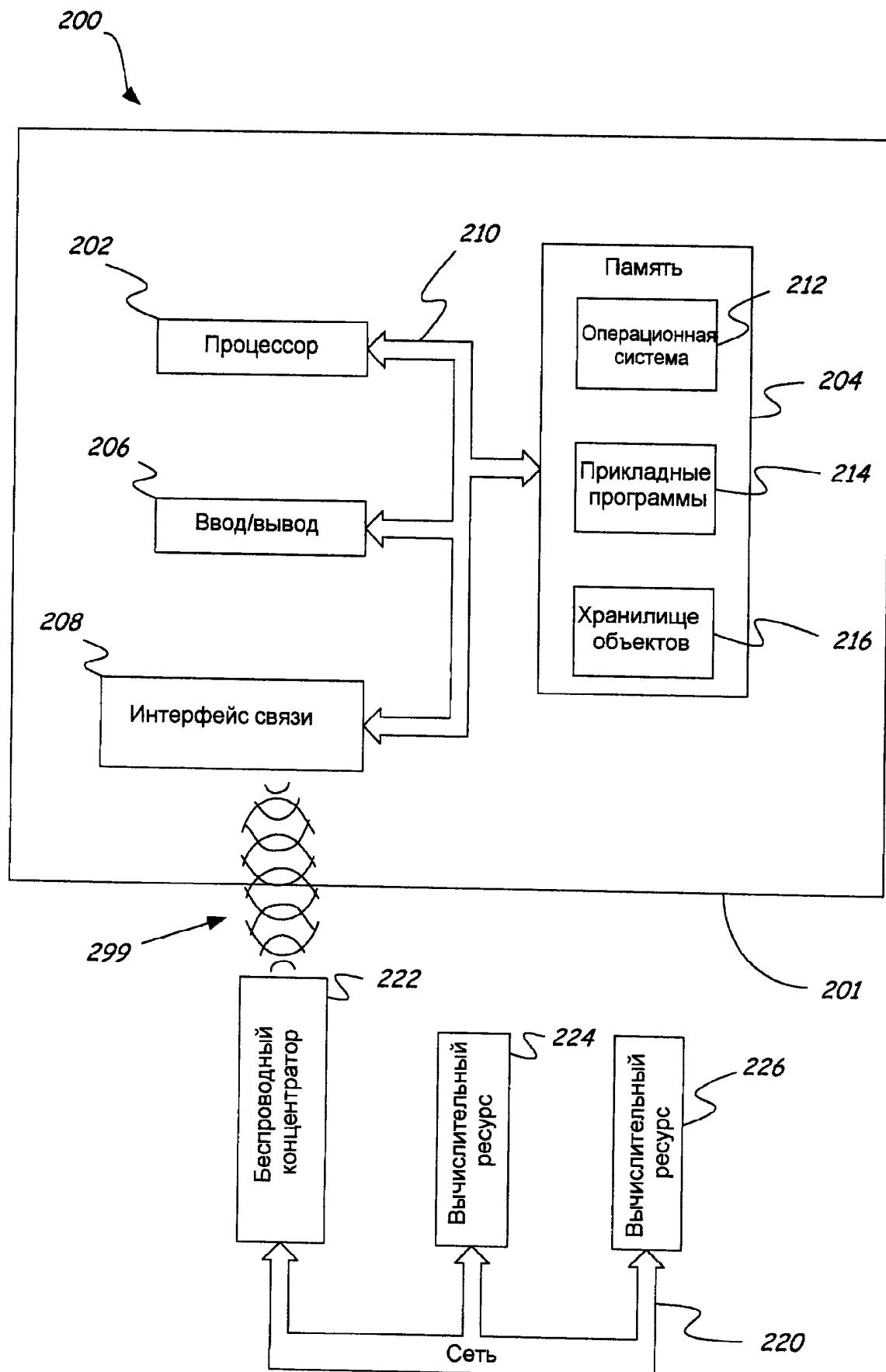
Фиг.2



Фиг.3



Фиг.5



Фиг.6