



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК
G06F 40/12 (2020.01)

(21)(22) Заявка: 2019139347, 02.12.2019

(24) Дата начала отсчета срока действия патента:
02.12.2019

Дата регистрации:
21.04.2020

Приоритет(ы):

(22) Дата подачи заявки: 02.12.2019

(45) Опубликовано: 21.04.2020 Бюл. № 12

Адрес для переписки:
197376, Санкт-Петербург, ул. Профессора
Попова, 5, СПбГЭТУ, патентный отдел,
Ивановой Е.А.

(72) Автор(ы):

Валов Александр Александрович (RU)

(73) Патентообладатель(и):

Федеральное государственное автономное
образовательное учреждение высшего
образования "Санкт-Петербургский
государственный электротехнический
университет "ЛЭТИ" им. В.И. Ульянова
(Ленина)" (RU)

(56) Список документов, цитированных в отчете
о поиске: US 2019/0155829 A1, 23.05.2019. US
10354009 B2, 16.07.2019. US 2018/0173698 A1,
21.06.2018. WO 2017/003889 A1, 05.01.2017. RU
2666277 C1, 06.09.2018.

(54) Способ содержательного анализа текстовой информации

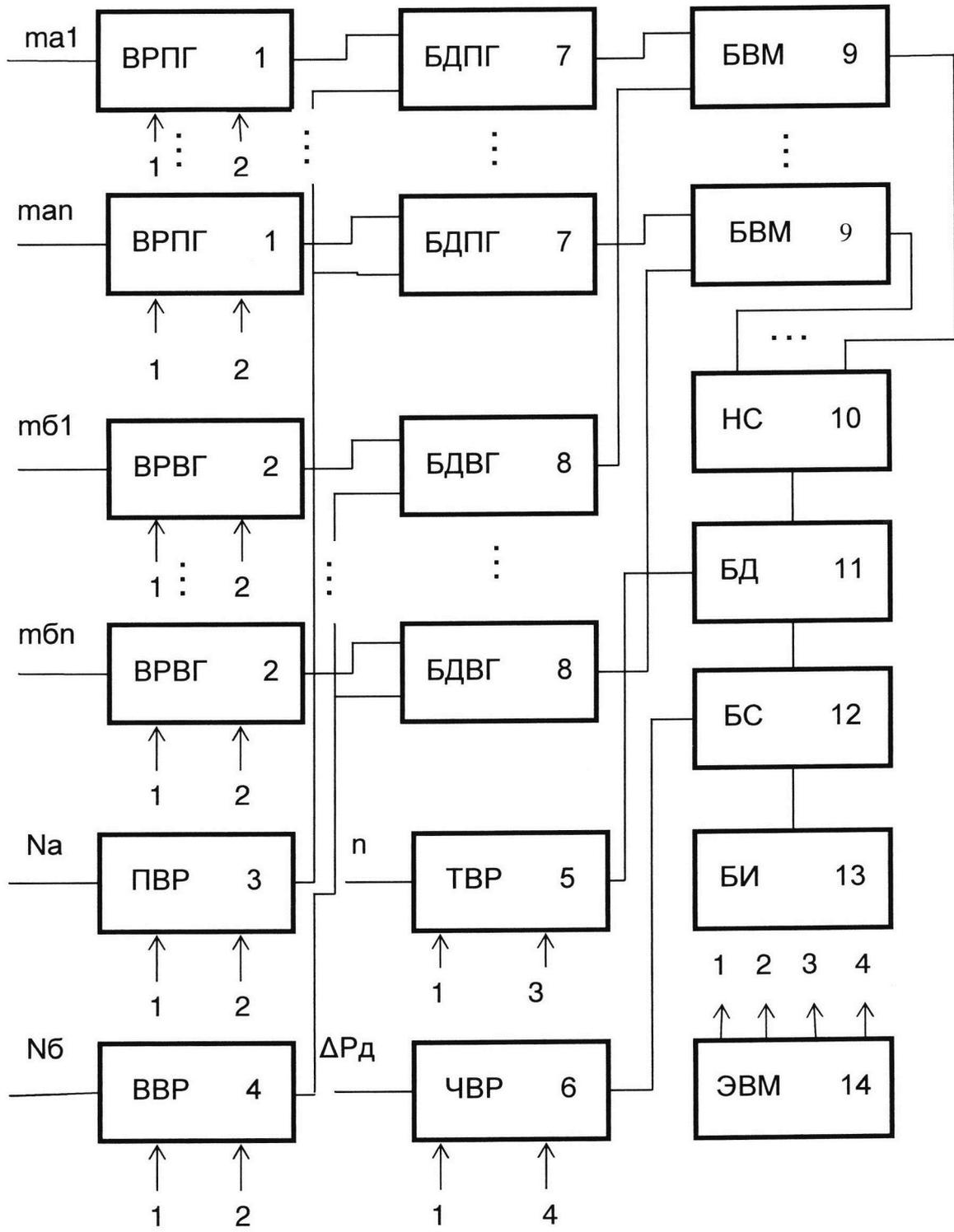
(57) Реферат:

Изобретение относится к области вычислительной техники. Технический результат заключается в повышении достоверности определения авторства текстовой информации. Технический результат достигается за счет расчленения первого и второго анализируемых текстов на нумеруемые фрагменты, для каждого из которых рассчитывают по известным формулам числовые значения пяти психолингвистических параметров. На основании подсчитанных значений параметров первого и второго текстов вычисляют с помощью метода факторного анализа соответственно первый и второй наборы нумеруемых значений общего фактора. Используя первый и второй

вычисленные наборы нумеруемых значений общего фактора, создают первую и вторую последовательности величин, каждая из которых представляет собой модуль разности между соседними числовыми элементами соответствующего набора значений общего фактора. Выделяют в первой и второй созданных числовых последовательностях соответственно первую и вторую совокупности равных чисел, при этом величину числа из первой выделенной совокупности обозначают $\Delta 1$, а величину числа из второй выделенной совокупности обозначают $\Delta 2$. Делают вывод о принадлежности первого и второго текстов одному автору, если $|\Delta 1 - \Delta 2| \leq 0.01$, а в противном случае авторы различны. 11 ил.

RU 2 7 1 9 5 5 3 C 1

RU 2 7 1 9 5 5 3 C 1



Фиг. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC
G06F 40/12 (2020.01)

(21)(22) Application: **2019139347, 02.12.2019**

(24) Effective date for property rights:
02.12.2019

Registration date:
21.04.2020

Priority:

(22) Date of filing: **02.12.2019**

(45) Date of publication: **21.04.2020** Bull. № 12

Mail address:

**197376, Sankt-Peterburg, ul. Professora Popova,
5, SPbGETU, patentnyj otdel, Ivanovoj E.A.**

(72) Inventor(s):

Valov Aleksandr Aleksandrovich (RU)

(73) Proprietor(s):

**Federalnoe gosudarstvennoe avtonomnoe
obrazovatelnoe uchrezhdenie vysshego
obrazovaniya "Sankt-Peterburgskij
gosudarstvennyj elektrotekhnicheskij universitet
"LETI" im. V.I. Ulyanova (Lenina)" (RU)**

(54) **METHOD OF SUBSTANTIVE ANALYSIS OF TEXT INFORMATION**

(57) Abstract:

FIELD: computer equipment.

SUBSTANCE: technical result is achieved due to the separation of the first and second analyzed texts into numbered fragments, for each of which based on known formulas numerical values of five psycholinguistic parameters are calculated. Based on the calculated values of the parameters of the first and second texts, the first and second sets of the numbered values of the common factor are calculated using the factor analysis method, respectively. Using the first and second calculated sets of numbered values of the common factor, the first and second sequences of values are created, each of which is a module of difference

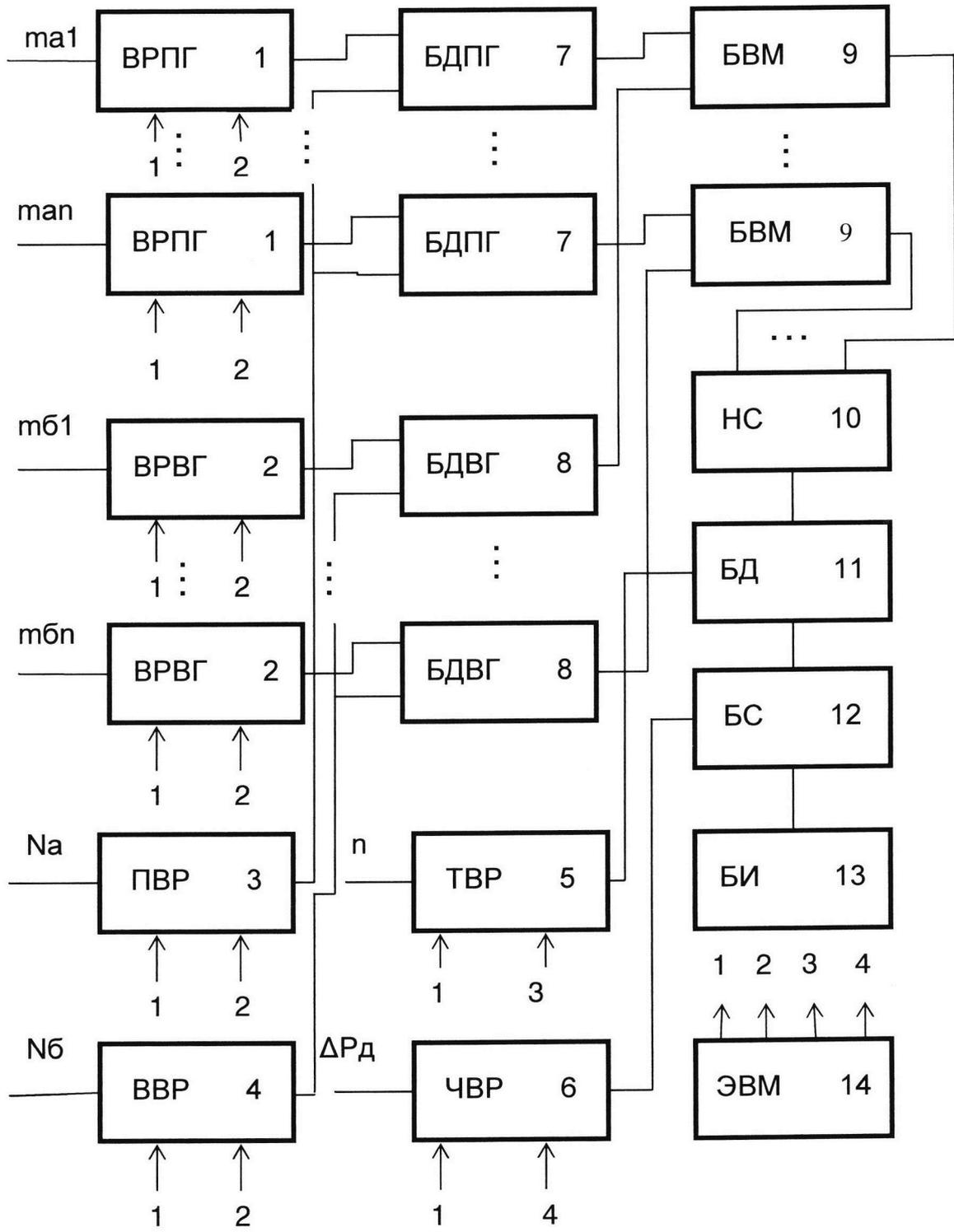
between adjacent numerical elements of the corresponding set of values of the common factor. First and second sets of equal numbers are selected in the first and second created numerical sequences, wherein the value of the number from the first selected collection is $\Delta 1$, and the value of the number from the second selected collection is $\Delta 2$. It is concluded that the first and second texts belong to one author, if $|\Delta 1 - \Delta 2| \leq 0.01$, and otherwise the authors are different.

EFFECT: technical result consists in improvement of authenticity of attribution of textual information.

1 cl, 11 dwg

RU 2 719 553 C1

RU 2 719 553 C1



Фиг. 1

Изобретение относится к вычислительной технике. Предлагаемый способ предназначен для идентификации текстовой информации в случаях определения ее автора.

Из информационных источников известно устройство, обеспечивающее техническое решение задачи определения авторства текстовой информации. Описание указанного устройства представлено в статье: алгоритм идентификации текстовой информации / А.В. Полтавский, Н.К. Юрков, А.В. Гриншкун // Надежность и качество сложных систем. - 2017. - №1 (17). - С. 77-84. DOI 10.21685/2307-4205-2017-1-10. Представленное в информационном источнике устройство осуществляет способ содержательного анализа текстовой информации.

Известен способ наиболее близкий по технической сущности к предлагаемому изобретению, описание которого представлено в патенте РФ №2568272, кл. G06F 17/00 2014.16.04. В известном способе содержательного анализа текстовой информации, позволяющем определять авторство текстовой информации, выполняют совокупность описанных ниже операций, действие которых в пространстве и времени инициируют управляющими сигналами, подаваемыми по соответствующим нумеруемым линиям связи, которые соединяют каждый из блоков (акторов), выполняющих соответствующую операцию, и управляющий блок, который своевременно формирует сигнал (импульс) и передает его по соответствующей нумеруемой линии для инициирования операции в акторе. Указанный способ выбран в качестве прототипа. В прототипе предусмотрено выполнение следующей совокупности операций.

Подсчитывают в каждом из двух анализируемых текстов «а» и «б» общее количество букв, обозначаемое соответственно N_a и N_b .

Генерируют начальный управляющий сигнал, который синхронизирует операцию сохранения подсчитанных величин N_a , N_b для их последующего использования, инициируемого вторым сигналом управления.

Подсчитывают количества каждой буквы алфавита в анализируемых текстах, при этом количество i -й буквы в тексте «а» обозначают m_{ai} , а количество j -й буквы в тексте «б» обозначают m_{bj} .

Сохраняют по начальному управляющему сигналу величины m_{ai} , m_{bj} для их последующего использования, инициируемого вторым управляющим сигналом.

Задают количество букв, имеющихся в используемом алфавите, которое обозначают n .

Сохраняют по начальному управляющему сигналу величину ΔP_d , для последующего использования, инициируемого третьим управляющим сигналом.

Задают фиксированную величину, обозначаемую ΔP_d , которую сохраняют по начальному управляющему сигналу для последующего использования, инициируемого четвертым управляющим сигналом.

Вычисляют по второму управляющему сигналу первый набор n величин и второй набор n величин, каждую из которых получают в результате выполнения операции деления согласно следующим формулам 1 и 2.

$$P_{ai} = m_{ai} / N_a, \quad (1)$$

где $i=1, \dots, n$.

$$P_{bj} = m_{bj} / N_b, \quad (2)$$

где $j=1, \dots, n$. При этом каждая величина, полученная в результате выполнения операции деления, характеризует вероятность появления соответствующей буквы в соответствующем тексте.

Определяют набор, состоящий из n величин, для получения каждой из которых вычисляют модуль разности между величиной вероятности появления конкретной буквы в тексте «а» и величиной вероятности появления этой же буквы в тексте «б».

Выполняют операцию суммирования определенных ранее n величин, каждая из которых соответствует модулю разности между величинами, соответствующими вероятностям появления конкретной буквы в текстах «а» и «б».

Выполняют по третьему управляющему сигналу операцию деления на величину n вычисленного ранее значения суммы величин модулей разностей между величиной вероятности появления конкретной буквы в тексте «а» и величиной вероятности появления этой же буквы в тексте «б» для получения среднего значения указанной разности согласно следующей формулы (3).

$$\Delta P_{\text{ср}} = \left(\sum_{k=1}^n |P_{ai} - P_{bj}| \right) / n, \quad (3)$$

где $k=1, \dots, n, i=1, \dots, n, j=1, \dots, n$.

Сравнивают по четвертому управляющему сигналу вычисленную величину $\Delta P_{\text{ср}}$ с допустимым фиксированным значением $\Delta P_{\text{д}}$ и при выполнении неравенства $\Delta P_{\text{ср}} \leq \Delta P_{\text{д}}$ вырабатывают сигнал оповещения о принадлежности двух анализируемых текстов «а» и «б» одному автору.

Указанный прототип имеет существенный недостаток, заключающийся в том, что способ-прототип обеспечивает весьма низкую достоверность определения авторства текстовой информации из-за использования только одного параметра анализируемого текста. В прототипе в качестве параметра используют среднее значение разности между вероятностью появления i -й буквы в первом анализируемом тексте и вероятностью появления i -й буквы во втором анализируемом тексте. Если значение параметра не превышает допустимой величины, то вырабатывают сигнал оповещения о принадлежности обоих анализируемых текстов одному автору.

Другим недостатком прототипа является использование самого нижнего уровня структурно-иерархической модели анализируемого текста (уровень букв). В исследовании А.С. Сурковой приводится утверждение о том, что на нижних уровнях (уровень букв) в большей степени проявляются универсальные инварианты текста. Вероятность появления отдельной буквы в тексте не зависит от текста и автора. На верхних уровнях (уровни предложений и слов) ярче проявляются авторские инварианты, индивидуальные особенности владения языком. Результаты исследования опубликованы в журнале Информационные технологии. Вестник Нижегородского университета им. Н.И. Лобачевского, 2014, №3 (1), с. 145-149. Идентификация авторства текстов на основе информационных портретов.

В задаче, на решение которой направлено заявляемое техническое решение, требуется создать способ, позволяющий идентифицировать текстовую информацию в случаях определения ее автора.

Технический результат заявляемого изобретения заключается в повышении достоверности определения авторства текстовой информации. Полученный технический результат обеспечивает возможность принятия объективных решений при защите авторских прав создателей текста и других подобных объектов, связанных с правом интеллектуальной собственности.

Указанный технический результат получают за счет того, что способ содержательного анализа текстовой информации, согласно которому подсчитывают в каждом из двух анализируемых текстов «а» и «б» общее количество букв, обозначаемое соответственно N_a и N_b , генерируют начальный управляющий сигнал, который синхронизирует

операцию сохранения подсчитанных величин N_a , N_b для их последующего использования, инициируемого вторым сигналом управления, подсчитывают количества каждой буквы алфавита в анализируемых текстах, при этом количество i -й буквы в тексте «а» обозначают mai , а количество j -й буквы в тексте «б» обозначают mbj , сохраняют по начальному управляющему сигналу величины mai , mbj для их последующего использования, инициируемого вторым управляющим сигналом, задают количество букв, имеющихся в используемом алфавите, которое обозначают n , сохраняют по начальному управляющему сигналу величину n для последующего использования, инициируемого третьим управляющим сигналом, задают фиксированную величину, обозначаемую ΔP_d , которую сохраняют по начальному управляющему сигналу для последующего использования, инициируемого четвертым управляющим сигналом, вычисляют по второму управляющему сигналу первый набор n величин и второй набор n величин, каждую из которых получают в результате выполнения операции деления согласно следующим формулам $P_{ai} = mai / N_a$, где $i = 1, \dots, n$, $P_{bj} = mbj / N_b$, где $j = 1, \dots, n$, при этом каждая величина, полученная в результате выполнения операции деления, характеризует вероятность появления соответствующей буквы в соответствующем тексте, определяют набор, состоящий из n величин, для получения каждой из которых вычисляют модуль разности между величиной вероятности появления конкретной буквы в тексте «а» и величиной вероятности появления этой же буквы в тексте «б», выполняют операцию суммирования определенных ранее n величин, каждая из которых соответствует модулю разности между величинами, соответствующими вероятностям появления конкретной буквы в текстах «а» и «б», выполняют по третьему управляющему сигналу операцию деления на величину n вычисленного ранее значения суммы величин модулей разностей между величиной вероятности появления конкретной буквы в тексте «а» и величиной вероятности появления этой же буквы в тексте «б» для получения среднего значения указанной разности согласно следующей формуле $\Delta P_{ср} = \left(\sum_{k=1}^n |P_{ai} - P_{bj}| \right) / n$, где $k = 1, \dots, n$, $i = 1, \dots, n$, $j = 1, \dots, n$, сравнивают по четвертому управляющему сигналу вычисленную величину $\Delta P_{ср}$ с допустимым фиксированным значением ΔP_d и при выполнении неравенства $\Delta P_{ср} \leq \Delta P_d$ вырабатывают сигнал оповещения о принадлежности двух анализируемых текстов «а» и «б» одному автору, дополнительно расчленяют первый и второй анализируемые тексты на нумеруемые фрагменты, при этом количество выделенных фрагментов первого текста равно количеству выделенных фрагментов второго текста, выделяют пять психолингвистических параметров, используемых при содержательном анализе текстовой информации, а именно процент служебных слов, содержащихся в анализируемом тексте ($\%сс$), коэффициент лексического разнообразия ($клр$), коэффициент логической связности ($клс$), среднюю длину слова ($Дс$), среднюю длину предложения ($Дп$), при этом подсчитывают для каждого выделенного фрагмента первого и второго текстов значение первого параметра по формуле $\%сс = ксс / K * 100$, где $ксс$ соответствует количеству служебных слов в анализируемом тексте, K соответствует количеству всех слов в тексте, подсчитывают значение второго параметра по формуле $клр = кспип / \sqrt{2K}$, где $кспип$ соответствует количеству слов в тексте после исключения повторяющихся, подсчитывают значение третьего параметра по формуле $клс = ксс / 3N$, где N соответствует количеству предложений в тексте, подсчитывают значение четвертого параметра по формуле $Дс = кб / K$, где $кб$ соответствует количеству букв в тексте, подсчитывают значение пятого

параметра по формуле $D_p = K/N$, на основании подсчитанных значений параметров первого и второго текстов вычисляют с помощью метода факторного анализа соответственно первый и второй наборы нумеруемых значений общего фактора, создают, используя первый и второй вычисленные наборы нумеруемых значений общего фактора, первую и вторую последовательности величин, каждая из которых представляет собой модуль разности между соседними числовыми элементами соответствующего набора значений общего фактора, выделяют в первой и второй созданных числовых последовательностях соответственно первую и вторую совокупности равных чисел, при этом величину числа из первой выделенной совокупности обозначают Δ_1 , а величину числа из второй выделенной совокупности обозначают Δ_2 , делают вывод о принадлежности первого и второго текстов одному автору, если $|\Delta_1 - \Delta_2| \leq 0.01$, а в противном случае авторы указанных текстов различны.

Сущность заявляемого изобретения поясняется на фигурах, где на фиг. 1 изображена функциональная схема устройства, реализующего способ содержательного анализа текстовой информации (с целью улучшения наглядности схемы связи между ЭВМ и управляющими входами соответствующих блоков показаны не полностью, а обозначены путем нумерации входов и выходов); на фиг. 2 изображена циклограмма работы устройства (на оси ординат обозначены номера выходов ЭВМ, а на оси абсцисс - число тактов), причем длительность различных вычислительных операций (сложение и вычитание - один такт, деление - двенадцать тактов) представлена в верхней части фиг. 2; на фиг. 3 изображена табл. 1, состоящая из пяти столбцов и двадцати строк, нумеруемых сверху вниз, в каждой из которых приведены числовые значения пяти психолингвистических параметров (%сс, клр, клс, Дс, Дп), рассчитанных для соответственно пронумерованных текстовых фрагментов, принадлежащих литературному произведению А.П. Чехова «Мужики» (<https://ilibrarv.ru/text/1160/P.1/index.html>), называемого для удобства изложения первым исследуемым текстом; на фиг. 4 изображена табл. 2, содержащая только один столбец, в котором приведен набор числовых значений фактора, вычисленных для исходных данных, представленных в табл. 1; на фиг. 5 изображен график 1 кусочно-линейной функции $Y=f(X)$, где Y числовое значение фактора, а X порядковый номер соответствующего анализируемого текстового фрагмента; на фиг. 6 изображена табл. 3, состоящая из пяти столбцов и двадцати строк, нумеруемых сверху вниз, в каждой из которых приведены числовые значения пяти психолингвистических параметров (%сс, клр, клс, Дс, Дп), рассчитанных для соответственно пронумерованных текстовых фрагментов, принадлежащих литературному произведению А.П. Чехова «Бабье царство» (<https://libkina.ru/books/prose-/prose-classic/169368-anton-chehov-babe-tsarstvo.html>), называемого для удобства изложения вторым исследуемым текстом; на фиг. 7 изображена табл. 4, содержащая только один столбец, в котором приведен набор числовых значений общего фактора, вычисленных для исходных данных, представленных в табл. 3; на фиг. 8 изображен график 2 кусочно-линейной функции $Y=f(X)$, где Y числовое значение фактора, а X порядковый номер соответствующего анализируемого текстового фрагмента; на фиг. 9 изображена табл. 5, состоящая из пяти столбцов и двадцати строк, нумеруемых сверху вниз, в каждой из которых приведены числовые значения пяти психолингвистических параметров (%сс, клр, клс, Дс, Дп), рассчитанных для соответственно пронумерованных текстовых фрагментов, принадлежащих литературному произведению В.Г. Короленко «История моего современника» (<https://ruslit.traumlibrarv.net/book/korolenko-ss10-05/korolenko-ss10-05.html>), называемого для удобства изложения третьим исследуемым текстом; на фиг. 10 изображена табл. 6, содержащая только один столбец, в котором приведен набор

числовых значений общего фактора, вычисленных для исходных данных, представленных в табл. 5; на фиг. 11 изображен график 3 кусочно-линейной функции $Y=f(X)$, где Y числовое значение общего фактора, а X порядковый номер соответствующего анализируемого текстового фрагмента.

5 Изображенные на фиг. 1 компоненты имеют буквенное обозначение и каждому компоненту сопоставлен номер. Номер 1 сопоставлен входным регистрам, составляющим первую группу и обозначаемым буквами ВРПГ. Номер 2 сопоставлен входным регистрам, составляющим вторую группу и обозначаемым буквами ВРВГ. Номера 3, 4, 5, 6 сопоставлены соответственно первому входному регистру (ПВР),
10 второму входному регистру (ВВР), третьему входному регистру (ТВР), четвертому входному регистру (ЧВР). Блокам деления первой группы (БДПГ) сопоставлен номер 7, а блокам деления второй группы (БДВГ) сопоставлен номер 8. Номер 9 сопоставлен блокам вычитания по модулю (БВМ). Номер 10 сопоставлен накопительному сумматору (НС), имеющему n входов. Номер 11 сопоставлен блоку деления (БД). Блоку сравнения
15 (БС) сопоставлен номер 12. Блоку индикации сопоставлен номер 13. Номер 14 сопоставлен ЭВМ.

Устройство, реализующее предлагаемый способ содержательного анализа текстовой информации (фиг. 1), содержит первую 1 и вторую 2 группы входных регистров, каждая из которых состоит из n элементов, первый 3, второй 4, третий 5 и четвертый 6 входные
20 регистры, первую 7 и вторую 8 группы блоков деления, каждая из которых состоит из n элементов, группу 9 блоков вычитания по модулю, состоящую из n элементов, накопительный сумматор 10, блок 11 деления, блок 12 сравнения, блок 13 индикации, ЭВМ 14, соответствующие выходы которой подключены к управляющим входам соответствующих блоков устройства.

25 Способ содержательного анализа текстовой информации осуществляется следующим образом. На информационные входы с первого по n -й элементов первой группы 1 входных регистров (фиг. 1) подаются соответственно величины $ma_1, \dots, ma_i, \dots, ma_n$, а на информационные входы с первого по n -й элементов второй группы 2 входных регистров подаются соответственно значения $mb_1, \dots, mb_i, \dots, mb_n$.

30 На информационные входы первого 3, второго 4, третьего 5 и четвертого 6 входных регистров направляются соответственно величины Na, Nb, n и ΔP_d , а управляющий сигнал на входы записи всех элементов этих групп входных регистров и входных регистров подается с первого выхода ЭВМ 14. При этом в качестве первого, второго, третьего и четвертого выходов ЭВМ используют соответствующие контакты разъема
35 порта LPT ЭВМ (пример дан в информационном источнике http://altay-krylov.ru/poleznaia_shemotehnika/proqrammirovanie_lpt_porta.html).

Управляющие сигналы, подаваемые на соответствующие контакты разъема порта LPT, формируют программно, с помощью таймера (пример дан в информационном источнике http://portal.tpu.ru/departments/kafedra/eafu/books_eafu1/Tab/LAB5.pdf).

40 По сигналу со второго выхода ЭВМ 14 на входы считывания первой 1 и второй 2 групп входных регистров величины ma_i и mb_j с их выходов засылаются на входы делимого соответственно первой 7 и второй 8 групп блоков деления. На входы делителя этих групп направляются по сигналу со второго выхода ЭВМ 14 с выходов соответственно первого 3 и второго 4 входных регистров значения Na и Nb . С выходов
45 первой 7 и второй 8 групп блоков деления величины Pa_i и Pb_j , определяемые по формулам (1) и (2), поступают соответственно на входы уменьшаемого и входы вычитаемого группы 9 блоков вычитания по модулю. С выходов этой группы величины $|Pa_i - Pb_j|_k$ засылаются на входы накопительного сумматора 10, с выхода которого

значение $\sum_{k=1}^n |P_{ai} - P_{bj}|_k$, подается на вход делимого блока 11 деления. На вход

5 делителя этого блока по сигналу с третьего выхода ЭВМ 14 направляется с выхода третьего входного регистра 5 величина n . С выхода блока 11 деления значение $\Delta P_{ср}$, определяемое по формуле (3), поступает на информационный вход блока 12 сравнения, на пороговый вход которого по сигналу с четвертого выхода ЭВМ 14 засылается с

10 выхода четвертого входного регистра 6 величина $\Delta P_{д}$. Если $\Delta P_{ср} \leq \Delta P_{д}$ (отрывки «а» и «б» принадлежат одному автору) на выходе блока 12 сравнения появится сигнал, который приведет к загоранию блока 13 индикации. В противном случае, когда $\Delta P_{ср} > \Delta P_{д}$ сигнала на выходе блока 12 сравнения не будет и блок 13 индикации не засветится, это будет свидетельствовать о том, что отрывки «а» и «б» принадлежат разным авторам. Порядок функционирования блоков устройства

15 представлен на циклограмме его работы (фиг. 2). Дополнительно расчленяют первый и второй анализируемые тексты на нумеруемые фрагменты, при этом количество выделенных фрагментов первого текста равно количеству выделенных фрагментов второго текста.

20 Выделяют пять психолингвистических показателей (параметров), используемых при содержательном анализе текстовой информации, а именно процент служебных слов, содержащихся в анализируемом тексте (%сс), коэффициент лексического разнообразия (клр), коэффициент логической связности (клс), среднюю длину слова (Дс), среднюю длину предложения (Дп).

25 Подсчитывают для каждого вычлененного фрагмента первого и второго текстов значение первого параметра по формуле $\%сс = ксс/К * 100$, где ксс соответствует количеству служебных слов в анализируемом тексте, К соответствует количеству всех слов в тексте.

Подсчитывают значение второго параметра по формуле $клр = кспип / \sqrt{2К}$, где кспип соответствует количеству слов в тексте после исключения повторяющихся.

30 Подсчитывают значение третьего параметра по формуле $клс = ксс/3N$, где N соответствует количеству предложений в тексте.

Подсчитывают значение четвертого параметра по формуле $Дс = кб/К$, где кб соответствует количеству букв в тексте.

Подсчитывают значение очередного пятого параметра по формуле $Дп = К/N$.

35 На основании подсчитанных значений параметров первого и второго текстов вычисляют с помощью метода факторного анализа соответственно первый и второй наборы нумеруемых значений общего фактора.

40 Создают, используя первый и второй вычисленные наборы нумеруемых значений общего фактора, первую и вторую последовательности величин, каждая из которых представляет собой модуль разности между соседними числовыми элементами соответствующего набора значений общего фактора.

Выделяют в первой и второй созданных числовых последовательностях соответственно первую и вторую совокупности равных чисел, при этом величину числа из первой выделенной совокупности обозначают $\Delta 1$, а величину числа из второй выделенной совокупности обозначают $\Delta 2$.

45 Делают вывод о принадлежности первого и второго текстов одному автору, если $|\Delta 1 - \Delta 2| \leq 0.01$, а в противном случае авторы указанных текстов различны.

Для подтверждения возможности определения авторства текстовой информации с помощью разработанного способа, целесообразно рассмотрение двух примеров

содержательного анализа текстовой информации. В первом примере предлагается использование двух литературных текстов, принадлежащих одному автору. Во втором примере предлагается использование двух литературных текстов, принадлежащих различным авторам.

5 В первом примере в качестве первого и второго исследуемых текстов предлагается использовать указанные выше произведения А.П. Чехова «Мужики» (1897 год) и «Бабье Царство» (1894 год). Во втором примере предлагается в качестве первого исследуемого текста использовать произведение А. П. Чехова «Мужики», а в качестве другого исследуемого текста использовать текст, соответствующий указанному выше
10 произведению В.Г. Короленко «История моего современника». При этом исследуемый текст В.Г. Короленко представлен предложениями, принадлежащими только первым пяти главам (первый том 1906 - 1909 годы) произведения.

Согласно предложенному изобретению в рассматриваемых примерах каждый исследуемый текст расчленяют на двадцать последовательно нумеруемых фрагментов.
15 Количество используемых текстовых фрагментов соответствует количеству наблюдений (измерений) при проведении факторного анализа. В информационном источнике (https://nafi.ru/upload/spss/Lecture_8.pdf) рекомендуют использовать количество наблюдений как минимум в два раза больше количества используемых переменных (параметров). Количество предложений, содержащихся в исследуемом текстовом фрагменте,
20 соответствует числу, размещенному внутри круглых скобок, перед которыми указан номер фрагмента.

Информация о количестве предложений, содержащихся во фрагментах, представлена следующим образом. Для первого текста: 1(15), 2(24), 3(26), 4(24), 5(17), 6(20), 7(35), 8(32), 9(33), 10(29), 11(41), 12(44), 13(42), 14(40), 15(38), 16(45), 17(27), 18(31), 19(21), 20(22).
25 Для второго текста: 1(34), 2(26), 3(33), 4(66), 5(51), 6(42), 7(35), 8(40), 9(37), 10(47), 11(41), 12(43), 13(33), 14(35), 15(33), 16(29), 17(39), 18(68), 19(26), 20(25). Для третьего текста: 1(26), 2(28), 3(27), 4(26), 5(20), 6(19), 7(19), 8(20), 9(20), 10(30), 11(22), 12(19), 13(18), 14(37), 15(22), 16(84), 17(34), 18(34), 19(39), 20(34).

Подсчет количества служебных слов в исследуемых предложениях осуществляется с помощью известного программного средства (<https://progaonline.com/syntax>). Указанное программное средство осуществляет пометку каждого служебного слова текста, предоставляемого программе для анализа. Пользовательский интерфейс указанного программного средства обеспечивает возможность прочтения на экране монитора ЭВМ размеченного текста и подсчета количества служебных слов в тексте. Количество
35 служебных слов, содержащихся в исследуемом текстовом фрагменте, представляется ниже числом, размещенным внутри круглых скобок, перед которыми указан номер исследуемого текстового фрагмента.

Результаты подсчета количества служебных слов в текстовых фрагментах представлены ниже для каждого из трех исследуемых текстов. Для первого текста: 1(53), 2(112), 3(26), 4(24), 5(62), 6(74), 7(162), 8(114), 9(114), 10(152), 11(160), 12(173), 13(160), 14(135), 15(143), 16(178), 17(188), 18(125), 19(118), 20(107). Для второго текста: 1(147), 2(173), 3(196), 4(165), 5(193), 6(176), 7(186), 8(186), 9(171), 10(174), 11(169), 12(173), 13(122), 14(116), 15(134), 16(204), 17(156), 18(165), 19(130), 20(137). Для третьего текста: 1(107), 2(130), 3(137), 4(129), 5(93), 6(84), 7(88), 8(86), 9(92), 10(119), 11(118), 12(91), 13(89),
45 14(156), 15(109), 16(316), 17(190), 18(103), 19(179), 20(159).

Подсчет в текстовых фрагментах количества всех слов и количества слов после исключения повторяющихся осуществляется с помощью известного программного средства (<http://wpcalc.com/kolichestvo-slov/>). Количество всех слов и количество слов

после исключения повторяющихся, содержащихся в исследуемом текстовом фрагменте, представляется ниже парой чисел, размещенных внутри круглых скобок, перед которыми указан номер исследуемого текстового фрагмента.

Результаты подсчета количества всех слов (К) и количества слов после исключения повторяющихся (ксипп) в текстовых фрагментах представлены ниже для каждого из трех исследуемых текстов. Для первого текста: 1(200, 157), 2(407, 296), 3(320, 237), 4(338, 248), 5(225, 169), 6(248, 185), 7(654, 423), 8(302, 216), 9(437, 303), 10(585, 380), 11(587, 398), 12(580, 396), 13(614, 417), 14(500, 347), 15(566, 386), 16(565, 382), 17(688, 408), 18(409, 273), 19(415, 291), 20(443, 322). Для второго текста: 1(620, 400), 2(674, 439), 3(654, 425), 4(682, 425), 5(739, 460), 6(640, 412), 7(686, 456), 8(663, 413), 9(624, 388), 10(604, 371), 11(647, 414), 12(671, 428), 13(489, 317), 14(540, 361), 15(495, 322), 16(659, 415), 17(553, 355), 18(644, 374), 19(486, 319), 20(450, 275). Для третьего текста: 1(417, 281), 2(490, 322), 3(497, 329), 4(456, 304), 5(383, 279), 6(330, 247), 7(352, 256), 8(309, 224), 9(352, 268), 10(543, 386), 11(427, 299), 12(366, 270), 13(346, 262), 14(600, 414), 15(455, 331), 16(1035, 588), 17(628, 397), 18(452, 328), 19(673, 432), 20(565, 387).

Подсчет количества букв в исследуемых текстовых фрагментах осуществляется с помощью текстового редактора Microsoft Word. Количество букв, содержащихся в исследуемом текстовом фрагменте, представляется ниже числом, размещенным внутри круглых скобок, перед которыми указан номер исследуемого текстового фрагмента.

Результаты подсчета количества букв в текстовых фрагментах представлены ниже для каждого из трех исследуемых текстов. Для первого текста: 1(1062), 2(2139), 3(1668), 4(1795), 5(1206), 6(1391), 7(3440), 8(1579), 9(2343), 10(2979), 11(3101), 12(3065), 13(3284), 14(2600), 15(3026), 16(2898), 17(3436), 18(2138), 19(2120), 20(2395). Для второго текста: 1(3433), 2(3532), 3(3366), 4(3566), 5(3811), 6(3287), 7(3559), 8(3439), 9(3236), 10(3066), 11(3445), 12(3485), 13(2568), 14(2911), 15(2662), 16(3337), 17(2887), 18(3245), 19(2549), 20(2249). Для третьего текста: 1(2201), 2(2633), 3(2703), 4(2447), 5(2121), 6(1937), 7(2017), 8(1747), 9(2104), 10(3235), 11(2296), 12(2121), 13(1976), 14(3310), 15(2612), 16(5203), 17(3486), 18(2457), 19(3538), 20(3159).

Для первого, второго и третьего исследуемых литературных текстов вычисляют значения пяти психолингвистических параметров, используя приведенные выше результаты подсчетов и соответствующие формулы. Вычисленные по формулам значения параметров размещают в соответствующих таблицах (табл. 1, 3, 5), номера строк которых соответствуют номерам анализируемых текстовых фрагментов.

В соответствии с демонстрационным примером 1 на основании подсчитанных значений параметров первого (табл. 1) и второго (табл. 3) текстов вычисляют с помощью метода факторного анализа соответственно первый и второй наборы нумеруемых значений общего фактора. Набор из двадцати значений общего фактора для первого текста представлен в табл. 2. Набор из двадцати значений общего фактора для второго текста представлен в табл. 4.

Создают, используя первый (табл. 2) и второй (табл. 4) вычисленные наборы нумеруемых значений общего фактора, первую и вторую последовательности величин, каждая из которых представляет собой модуль разности между соседними числовыми элементами соответствующего набора значений общего фактора.

Первая последовательность содержит следующие величины: 0.18622, 0.18621, 0.18622, 0.18622, 0.18622, 0.15135, 0.13763, 0.65305, 0.27061, 0.18622, 0.18622, 0.18622, 0.18622, 0.18622, 0.20999, 0.54941, 1.11573, 1.41168, 0.18621. Вторая последовательность содержит следующие величины: 0.30038, 0.10799, 0.33490, 0.22547, 0.18801, 0.18801, 0.18800, 0.18801, 0.18801, 0.14326, 0.47195, 0.18800, 0.00128, 0.32997, 0.18801, 0.04861, 0.22002, 0.61492,

0.28265.

Выделяют в первой и второй созданных числовых последовательностях соответственно первую и вторую совокупности равных чисел, при этом величину числа из первой выделенной совокупности обозначают $\Delta 1$, а величину числа из второй выделенной совокупности обозначают $\Delta 2$. В результате выделения первой и второй совокупностей равных величин получают $\Delta 1=0.1862$ и $\Delta 2=0.1880$. При этом в первой совокупности содержится 11 равных элементов, а во второй совокупности содержится 7 равных элементов. Числовое значение модуля разности величин $\Delta 1$ и $\Delta 2$ равно 0.0018.

Делают вывод о принадлежности первого и второго текстов одному автору, так как $|\Delta 1-\Delta 2|\leq 0.01$.

Ниже излагаются дополнительные пояснения процесса осуществления способа содержательного анализа текстовой информации. Для обеспечения наглядности и удобства создания последовательностей величин, указанных выше, и выделения среди созданных последовательностей равных величин, обозначаемых $\Delta 1$ и $\Delta 2$, целесообразно использовать графическое представление функции $Y=f(X)$, где Y является числовым значением общего фактора, зависящим от порядкового номера X соответствующего текстового фрагмента.

Строят с помощью программы EXCEL кусочно-линейные графики функции $Y=f(X)$ для двух созданных наборов значений общего фактора. График 1 представлен на фиг. 5, а график 2 представлен на фиг. 8.

На графике 1 выявляют отрезки, имеющие одинаковый отрицательный угол наклона. На графике 1 концам (точкам) таких отрезков соответствуют абсциссы $X_{1_2_3_4_5_6}$, $10_11_12_13_14_15$, 19_20 . Для выявленных отрезков рассчитывают значение углового коэффициента по формуле $|Y_i-Y_j|$, где $Y_i=f(X_i)$, $Y_j=f(X_j)$, $X_j>X_i$ и $X_j-X_i=1$, а упорядоченные пары (X_i, Y_i) и (X_j, Y_j) являются графиком, изображаемым в виде отрезка. Значения Y общего фактора представлены в табл. 2 и на графике 1.

Рассчитанное значение углового коэффициента представляет собой величину $\Delta 1$, значение которой равно 0.1862.

На графике 2 выявляют отрезки, имеющие одинаковый отрицательный угол наклона. На графике 2 концам (точкам) таких отрезков соответствуют абсциссы $X_{5_6_7_8_9_10}$, 12_13 , 15_16 . Для выявленных отрезков рассчитывают значение углового коэффициента по формуле $|Y_i-Y_j|$, где $Y_i=f(X_i)$, $Y_j=f(X_j)$, $X_j>X_i$ и $X_j-X_i=1$, а упорядоченные пары (X_i, Y_i) и (X_j, Y_j) являются графиком, изображаемым в виде отрезка. Значения Y общего фактора представлены в табл. 4 и на графике 2.

Рассчитанное значение углового коэффициента представляет собой величину $\Delta 2$, значение которой равно 0.1880.

Вычисляют модуль разности величин $\Delta 1$ и $\Delta 2$. Вычисленная величина $|\Delta 1-\Delta 2|=0.0018$ меньше величины 0.01, поэтому делают вывод о принадлежности первого (А.П. Чехов) и второго (А.П. Чехов) текстов одному автору.

В соответствии с демонстрационным примером 2 в качестве исследуемых текстов используют первый и третий тексты. При этом анализ первого текста не требуется, а используются готовые результаты расчета величины $\Delta 1$ (0.1862), полученные при выполнении примера 1. На основании подсчитанных значений параметров третьего (табл. 5) текста вычисляют с помощью метода факторного анализа для третьего исследуемого текста набор нумерованных значений общего фактора. Набор из двадцати значений общего фактора для третьего текста представлен в табл. 6.

Создают, используя вычисленный набор нумеруемых значений общего фактора,

последовательность величин, каждая из которых представляет собой модуль разности между соседними числовыми элементами вычисленного набора значений общего фактора.

5 Последовательность содержит следующие величины: 0.28001, 0.13234, 0.13235, 0.37305, 0.61476, 0.13235, 0.13234, 0.13235, 0.13234, 0.89011, 1.10584, 0.13234, 0.34611, 0.80845, 0.13234, 0.13235, 0.79877, 0.13235, 0.13234.

10 Выделяют в созданной числовой последовательности совокупность равных чисел, при этом величину числа из выделенной совокупности обозначают $\Delta 3$. В результате выделения совокупности равных величин получают $\Delta 3=0.1323$. При этом в выделенной совокупности содержится 11 равных элементов. Числовое значение модуля разности величин $\Delta 1$ и $\Delta 3$ равно 0.0539.

Делают вывод о принадлежности первого и третьего текстов различным авторам, так как неравенство $|\Delta 1-\Delta 3|\leq 0.01$ не выполняется.

15 Ниже излагаются пояснения процесса осуществления способа содержательного анализа текстовой информации с использованием графического представления функции $Y=f(X)$. Строят с помощью программы EXCEL кусочно-линейный график 3 для созданного набора значений общего фактора. График 3 представлен на фиг.11.

20 На графике 3 выявляют отрезки, имеющие одинаковый отрицательный угол наклона. На графике 3 концам (точкам) таких отрезков соответствуют абсциссы $X_{2_3_4}$, $X_{6_7_8_9_10}$, X_{12_13} , $X_{15_16_17}$, $X_{18_19_20}$. Для выявленных отрезков рассчитывают значение углового коэффициента по формуле $|Y_i - Y_j|$, где $Y_i=f(X_i)$, $Y_j=f(X_j)$, $X_j>X_i$ и $X_j-X_i=1$, а упорядоченные пары (X_i, Y_i) и (X_j, Y_j) являются графиком, изображаемым в виде отрезка. Значения Y общего фактора представлены в табл. 6 и на графике 3.

25 Рассчитанное значение углового коэффициента представляет собой величину $\Delta 3$, значение которой равно 0.1323.

30 Вычисляют модуль разности величин $\Delta 1$ и $\Delta 3$. Вычисленная величина $|\Delta 1-\Delta 3|=0.0539$ больше величины 0.01, поэтому делают вывод о принадлежности первого (А.П. Чехов) и третьего (В.Г. Короленко) текстов различным авторам. Неравенство $|\Delta 1-\Delta 3|\leq 0.01$ не выполняется.

Продемонстрированная последовательность действий для определения авторства текстовой информации доказывает возможность осуществления заявляемого способа.

35 Таким образом, последовательность действий, представляющих заявляемый способ содержательного анализа текстовой информации, позволяет идентифицировать текстовую информацию в случаях определения ее авторства.

40 Следовательно, решена задача создания способа, позволяющего идентифицировать текстовую информацию в случаях определения ее авторства, и получен технический результат, заключающийся в повышении достоверности определения авторства текстовой информации. Полученный технический результат обеспечивает возможность принятия объективных решений при защите авторских прав создателей текста и других подобных объектов, связанных с правом интеллектуальной собственности.

(57) Формула изобретения

45 Способ содержательного анализа текстовой информации, согласно которому подсчитывают в каждом из двух анализируемых текстов «а» и «б» общее количество букв, обозначаемое соответственно N_a и N_b , генерируют начальный управляющий сигнал, который синхронизирует операцию сохранения подсчитанных величин N_a , N_b для их последующего использования, инициируемого вторым сигналом управления,

подсчитывают количества каждой буквы алфавита в анализируемых текстах, при этом количество i -й буквы в тексте «а» обозначают mai , а количество j -й буквы в тексте «б» обозначают mbj , сохраняют по начальному управляющему сигналу величины mai , mbj для их последующего использования, инициируемого вторым управляющим сигналом, задают количество букв, имеющихся в используемом алфавите, которое обозначают n , сохраняют по начальному управляющему сигналу величину n для последующего использования, инициируемого третьим управляющим сигналом, задают фиксированную величину, обозначаемую ΔP_d , которую сохраняют по начальному управляющему сигналу для последующего использования, инициируемого четвертым управляющим сигналом, вычисляют по второму управляющему сигналу первый набор n величин и второй набор n величин, каждую из которых получают в результате выполнения операции деления согласно следующим формулам $P_{ai} = mai / N_a$, где $i = 1, \dots, n$, $P_{bj} = mbj / N_b$, где $j = 1, \dots, n$, при этом каждая величина, полученная в результате выполнения операции деления, характеризует вероятность появления соответствующей буквы в соответствующем тексте, определяют набор, состоящий из n величин, для получения каждой из которых вычисляют модуль разности между величиной вероятности появления конкретной буквы в тексте «а» и величиной вероятности появления этой же буквы в тексте «б», выполняют операцию суммирования определенных ранее n величин, каждая из которых соответствует модулю разности между величинами, соответствующими вероятностям появления конкретной буквы в текстах «а» и «б», выполняют по третьему управляющему сигналу операцию деления на величину n вычисленного ранее значения суммы величин модулей разностей между величиной вероятности появления конкретной буквы в тексте «а» и величиной вероятности появления этой же буквы в тексте «б» для получения среднего значения указанной разности согласно следующей формуле $\Delta P_{cp} = \left(\sum_{k=1}^n |P_{ai} - P_{bj}|_k \right) / n$, где $k = 1, \dots, n$, $i = 1, \dots, n$, $j = 1, \dots, n$, сравнивают по четвертому управляющему сигналу вычисленную величину ΔP_{cp} с допустимым фиксированным значением ΔP_d и при выполнении неравенства $\Delta P_{cp} \leq \Delta P_d$ вырабатывают сигнал оповещения о принадлежности двух анализируемых текстов «а» и «б» одному автору, отличающийся тем, что дополнительно расчленяют первый и второй анализируемые тексты на нумеруемые фрагменты, при этом количество выделенных фрагментов первого текста равно количеству выделенных фрагментов второго текста, выделяют пять психолингвистических параметров, используемых при содержательном анализе текстовой информации, а именно процент служебных слов, содержащихся в анализируемом тексте ($\%cc$), коэффициент лексического разнообразия ($клр$), коэффициент логической связности ($клс$), среднюю длину слова ($Дс$), среднюю длину предложения ($Дп$), при этом подсчитывают для каждого вычлененного фрагмента первого и второго текстов значение первого параметра по формуле $\%cc = k_{cc} / K \cdot 100$, где k_{cc} соответствует количеству служебных слов в анализируемом тексте, K соответствует количеству всех слов в тексте, подсчитывают значение второго параметра по формуле $клр = k_{спип} / \sqrt{2K}$, где $k_{спип}$ соответствует количеству слов в тексте после исключения повторяющихся, подсчитывают значение третьего параметра по формуле $клс = k_{cc} / 3N$, где N соответствует количеству предложений в тексте, подсчитывают значение четвертого параметра по формуле $Дс = kb / K$, где kb соответствует количеству букв в тексте, подсчитывают значение пятого параметра по формуле $Дп = K / N$, на основании подсчитанных значений параметров первого и второго текстов вычисляют

с помощью метода факторного анализа соответственно первый и второй наборы нумеруемых значений общего фактора, создают, используя первый и второй вычисленные наборы нумеруемых значений общего фактора, первую и вторую последовательности величин, каждая из которых представляет собой модуль разности между соседними числовыми элементами соответствующего набора значений общего фактора, выделяют в первой и второй созданных числовых последовательностях соответственно первую и вторую совокупности равных чисел, при этом величину числа из первой выделенной совокупности обозначают $\Delta 1$, а величину числа из второй выделенной совокупности обозначают $\Delta 2$, делают вывод о принадлежности первого и второго текстов одному автору, если $|\Delta 1 - \Delta 2| \leq 0.01$, а в противном случае авторы указанных текстов различны.

15

20

25

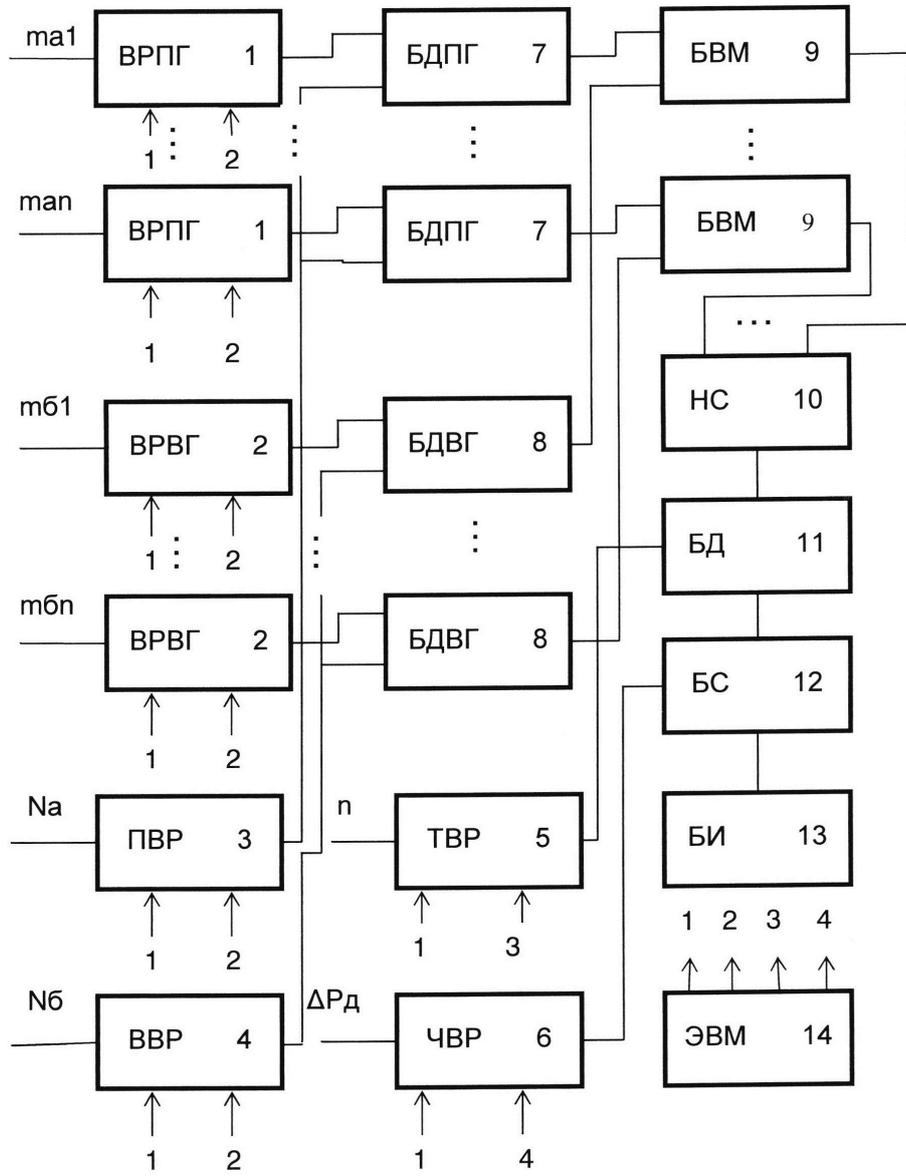
30

35

40

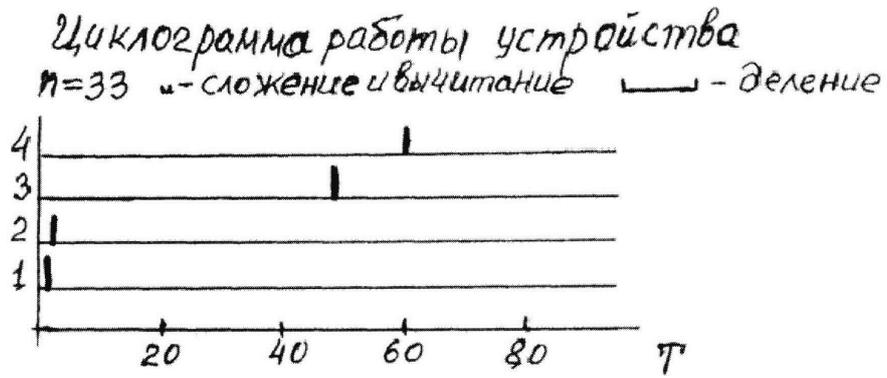
45

1



Фиг. 1

2



Фиг. 2

Таблица 1

№	%СС	КЛР (у. е.)	КЛС (у. е.)	Дс (букв)	Дп (слов)
1	2.650E+01	0.785E+01	1.178E+00	5.310E+00	1.333E+01
2	2.752E+01	1.037E+01	1.556E+00	5.256E+00	1.696E+01
3	2.813E+01	0.937E+01	1.154E+00	5.213E+00	1.231E+01
4	2.811E+01	0.954E+01	1.319E+00	5.311E+00	1.408E+01
5	2.756E+01	0.797E+01	1.216E+00	5.360E+00	1.324E+01
6	2.984E+01	0.831E+01	1.233E+00	5.609E+00	1.240E+01
7	2.477E+01	1.170E+01	1.543E+00	5.260E+00	1.869E+01
8	3.775E+01	0.879E+01	1.188E+00	5.228E+00	0.944E+01
9	2.609E+01	1.025E+01	1.152E+00	5.362E+00	1.324E+01
10	2.598E+01	1.111E+01	1.747E+00	5.092E+00	2.017E+01
11	2.726E+01	1.162E+01	1.301E+00	5.283E+00	1.432E+01
12	2.983E+01	1.163E+01	1.311E+00	5.284E+00	1.318E+01
13	2.606E+01	1.190E+01	1.270E+00	5.349E+00	1.462E+01
14	2.700E+01	1.097E+01	1.125E+00	5.200E+00	1.250E+01
15	2.526E+01	1.147E+01	1.254E+00	5.346E+00	1.489E+01
16	3.150E+01	1.136E+01	1.319E+00	5.129E+00	1.256E+01
17	2.732E+01	1.100E+01	2.321E+00	4.994E+00	2.548E+01
18	3.056E+01	0.954E+01	1.344E+00	5.227E+00	1.319E+01
19	2.843E+01	1.010E+01	1.873E+00	5.108E+00	1.976E+01
20	2.415E+01	1.082E+01	1.621E+00	5.406E+00	2.014E+01

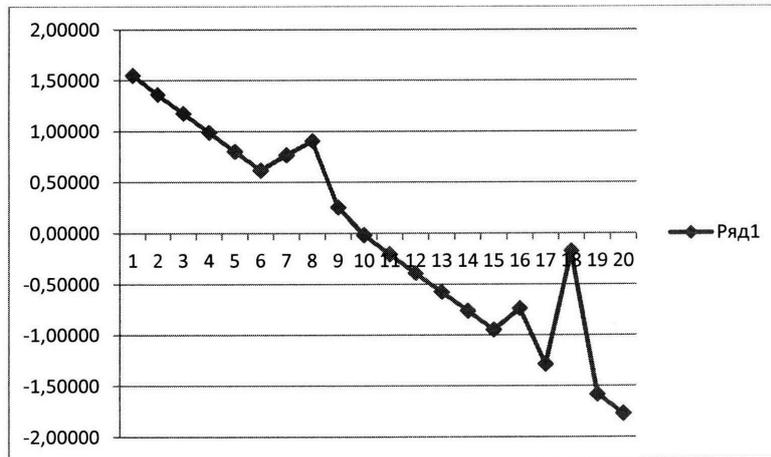
Фиг. 3

Таблица 2

№	Фактор
1	1,54963
2	1,36341
3	1,17720
4	0,99098
5	0,80476
6	0,61854
7	0,76989
8	0,90752
9	0,25447
10	-0,01614
11	-0,20236
12	-0,38858
13	-0,57480
14	-0,76102
15	-0,94724
16	-0,73725
17	-1,28666
18	-0,17093
19	-1,58261
20	-1,76882

Фиг. 4

График 1 функции $Y = f(X)$



Фиг. 5

Таблица 3

№	%СС	КЛР (у. е.)	КЛС (у. е.)	Дс (букв)	Дп (слов)
1	2.371E+01	1.135E+01	1.441E+00	5.537E+00	1.823E+01
2	2.566E+01	1.195E+01	2.217E+00	5.240E+00	2.592E+01
3	2.996E+01	1.175E+01	1.979E+00	5.146E+00	1.981E+01
4	2.419E+01	1.150E+01	0.833E+00	5.228E+00	1.033E+01
5	2.611E+01	1.196E+01	1.261E+00	5.156E+00	1.449E+01
6	2.750E+01	1.151E+01	1.396E+00	5.135E+00	1.523E+01
7	2.711E+01	1.231E+01	1.771E+00	5.188E+00	1.960E+01
8	2.805E+01	1.134E+01	1.550E+00	5.187E+00	1.657E+01
9	2.740E+01	1.098E+01	1.540E+00	5.185E+00	1.686E+01
10	2.880E+01	1.067E+01	1.234E+00	5.076E+00	1.285E+01
11	2.610E+01	1.150E+01	1.373E+00	5.324E+00	1.578E+01
12	2.578E+01	1.168E+01	1.341E+00	5.193E+00	1.560E+01
13	2.494E+01	1.013E+01	1.232E+00	5.251E+00	1.481E+01
14	2.148E+01	1.098E+01	1.104E+00	5.390E+00	1.542E+01
15	2.707E+01	1.023E+01	1.353E+00	5.377E+00	1.500E+01
16	3.095E+01	1.143E+01	1.344E+00	5.063E+00	2.272E+01
17	2.820E+01	1.067E+01	1.333E+00	5.220E+00	1.417E+01
18	2.562E+01	1.042E+01	0.808E+00	5.038E+00	0.947E+01
19	2.674E+01	1.023E+01	1.666E+00	5.244E+00	1.869E+01
20	3.044E+01	0.916E+01	1.826E+00	4.997E+00	1.800E+01

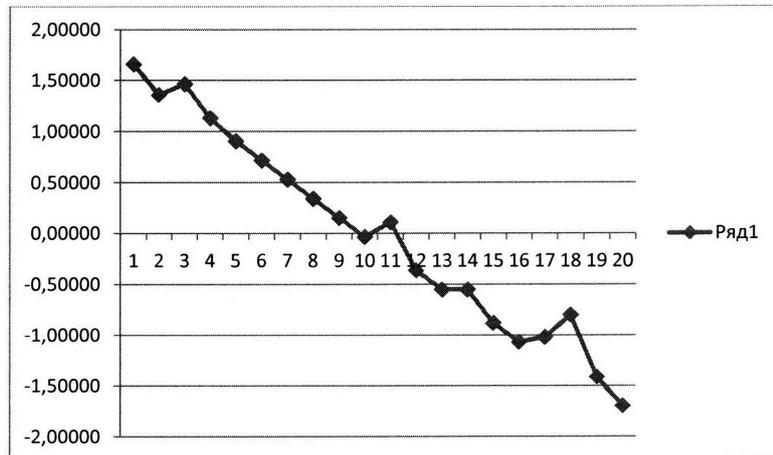
Фиг. 6

Таблица 4

№	Фактор
1	1,65863
2	1,35825
3	1,46624
4	1,13134
5	0,90587
6	0,71786
7	0,52985
8	0,34185
9	0,15384
10	-0,03417
11	0,10909
12	-0,36286
13	-0,55086
14	-0,54958
15	-0,87955
16	-1,06756
17	-1,01895
18	-0,79893
19	-1,41385
20	-1,69650

Фиг. 7

График 2 функции $Y = f(X)$



Фиг. 8

Таблица 5

№	%СС	КЛР (у. е.)	КЛС (у. е.)	Дс (букв)	Дп (слов)
1	2.566E+01	0.973E+01	1.372E+00	5.278E+00	1.604E+01
2	2.653E+01	1.025E+01	1.548E+00	5.373E+00	1.750E+01
3	2.757E+01	1.044E+01	1.691E+00	5.439E+00	1.841E+01
4	2.829E+01	1.007E+01	1.654E+00	5.366E+00	1.754E+01
5	2.428E+01	1.008E+01	1.550E+00	5.538E+00	1.915E+01
6	2.545E+01	0.961E+01	1.474E+00	5.870E+00	1.737E+01
7	2.500E+01	0.965E+01	1.544E+00	5.730E+00	1.853E+01
8	2.783E+01	0.901E+01	1.433E+00	5.654E+00	1.545E+01
9	2.614E+01	1.010E+01	1.533E+00	5.977E+00	1.760E+01
10	2.192E+01	1.171E+01	1.322E+00	5.958E+00	1.810E+01
11	2.763E+01	1.023E+01	1.788E+00	5.377E+00	1.941E+01
12	2.486E+01	0.998E+01	1.596E+00	5.795E+00	1.926E+01
13	2.572E+01	0.996E+01	1.648E+00	5.711E+00	1.922E+01
14	2.600E+01	1.195E+01	1.405E+00	5.517E+00	1.622E+01
15	2.396E+01	1.097E+01	1.652E+00	5.741E+00	2.068E+01
16	3.053E+01	1.292E+01	1.254E+00	5.027E+00	1.232E+01
17	3.025E+01	1.120E+01	1.863E+00	5.551E+00	1.847E+01
18	2.279E+01	1.091E+01	1.010E+00	5.436E+00	1.329E+01
19	2.660E+01	1.178E+01	1.530E+00	5.257E+00	1.726E+01
20	2.814E+01	1.206E+01	1.559E+00	5.591E+00	1.662E+01

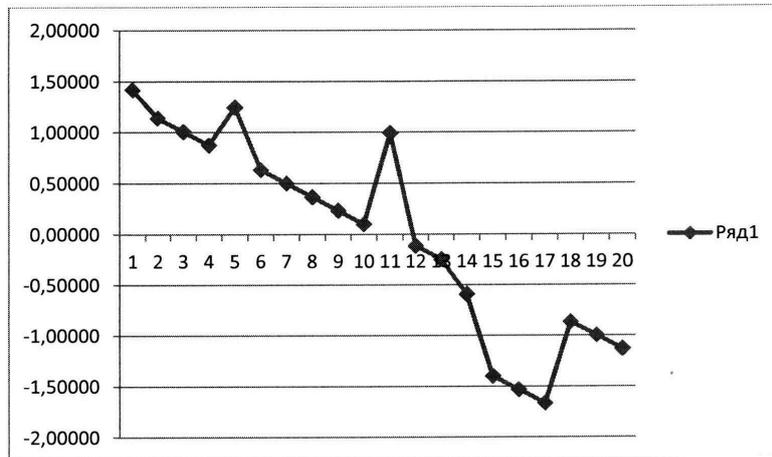
Фиг. 9

Таблица 6

№	Фактор
1	1,41988
2	1,13987
3	1,00753
4	0,87518
5	1,24823
6	0,63347
7	0,50112
8	0,36878
9	0,23643
10	0,10409
11	0,99420
12	-0,11164
13	-0,24398
14	-0,59009
15	-1,39854
16	-1,53088
17	-1,66323
18	-0,86446
19	-0,99681
20	-1,12915

Фиг. 10

График 3 функции $Y = f(X)$



Фиг. 11