



**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: 2016136531, 12.09.2016

(24) Дата начала отсчета срока действия патента:
12.09.2016Дата регистрации:
08.12.2017

Приоритет(ы):

(22) Дата подачи заявки: 12.09.2016

(45) Опубликовано: 08.12.2017 Бюл. № 34

Адрес для переписки:

119021, Москва, ул. Льва Толстого, 16, ООО
"ЯНДЕКС", патентный отдел

(72) Автор(ы):

САДОВСКИЙ Александр Анатольевич
(RU)

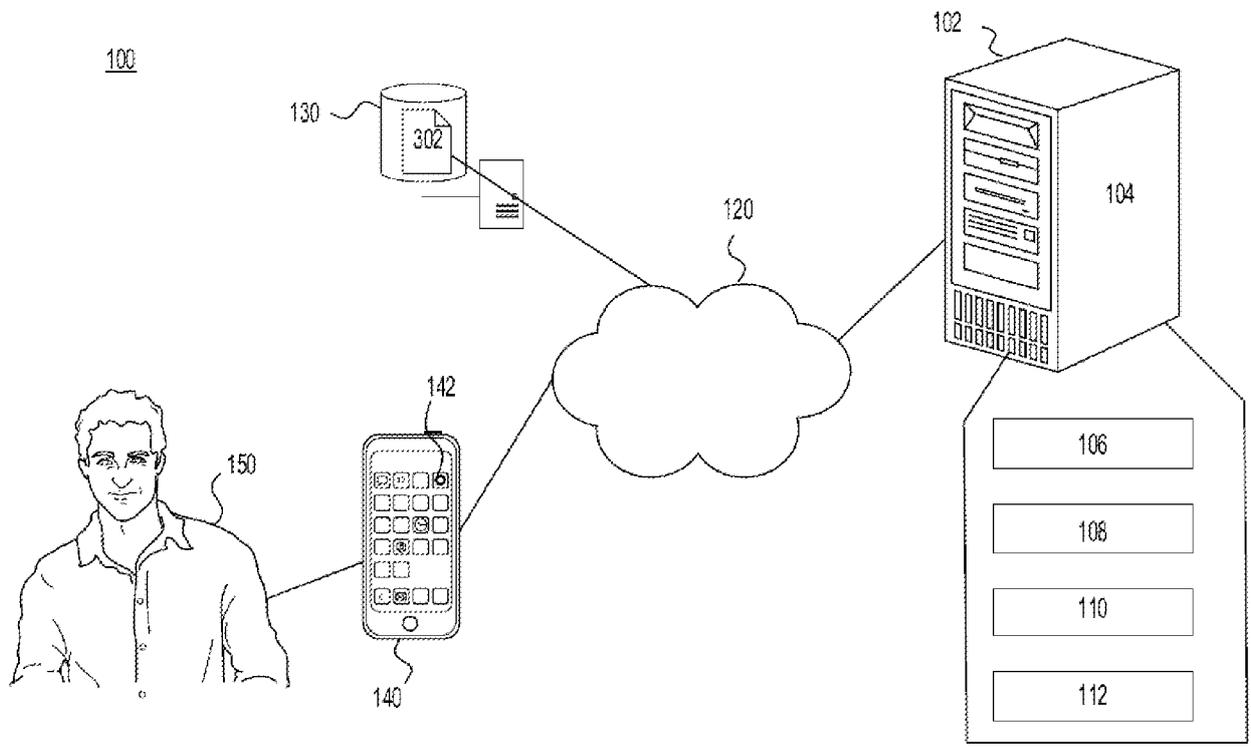
(73) Патентообладатель(и):

Общество с ограниченной ответственностью
"Яндекс" (RU)(56) Список документов, цитированных в отчете
о поиске: US 2014/0156651 A1, 05.06.2014. US
2014/0072285 A1, 13.03.2014. US 2015/0279390
A1, 01.10.2015. US 2005/0091203 A1, 28.04.2005.
US 2002/0175932 A1, 28.11.2002. RU 2440606
С2, 20.01.2012.**(54) СПОСОБ И СИСТЕМА СОЗДАНИЯ КРАТКОГО ИЗЛОЖЕНИЯ ЦИФРОВОГО КОНТЕНТА**

(57) Реферат:

Изобретение относится к обработке цифрового контента (в частности, текстовых, аудио- и видеофайлов), а конкретнее к созданию кратких изложений цифрового контента. Техническим результатом является расширение арсенала средств создания кратких изложений цифрового контента. В способе создания краткого изложения цифрового контента получают указание на цифровой контент и выполняют синтаксический анализ текстового представления контента. Разделяют контент на упорядоченное множество фрагментов, включающее в себя первый и второй фрагменты. Выполняют семантический анализ каждого

фрагмента и определяют параметр полезности для каждого фрагмента и связи между каждой парой фрагментов. В ответ на то, что параметр полезности второго фрагмента превышает предварительно определенное пороговое значение, включают второй фрагмент в подмножество фрагментов для включения в краткое изложение цифрового контента. В ответ на получение указания на связь второго фрагмента с первым включают первый фрагмент в подмножество фрагментов. Создают краткое изложение цифрового контента на основе подмножества фрагментов. 2 н. и 60 з.п. ф-лы, 6 ил.



ФИГ. 1

RU 2637998 C1

RU 2637998 C1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.
G06F 17/27 (2006.01)
G06F 17/30 (2006.01)

(12) **ABSTRACT OF INVENTION**(21)(22) Application: **2016136531, 12.09.2016**(24) Effective date for property rights:
12.09.2016Registration date:
08.12.2017

Priority:

(22) Date of filing: **12.09.2016**(45) Date of publication: **08.12.2017** Bull. № 34

Mail address:

**119021, Moskva, ul. Lva Tolstogo, 16, OOO
"YANDEKS", patentnyj otdel**

(72) Inventor(s):

SADOVSKIJ Aleksandr Anatolevich (RU)

(73) Proprietor(s):

**Obshchestvo s ogranichennoj otvetstvennostyu
"Yandeks" (RU)**(54) **METHOD AND SYSTEM FOR CREATING BRIEF SUMMARY OF DIGITAL CONTENT**

(57) Abstract:

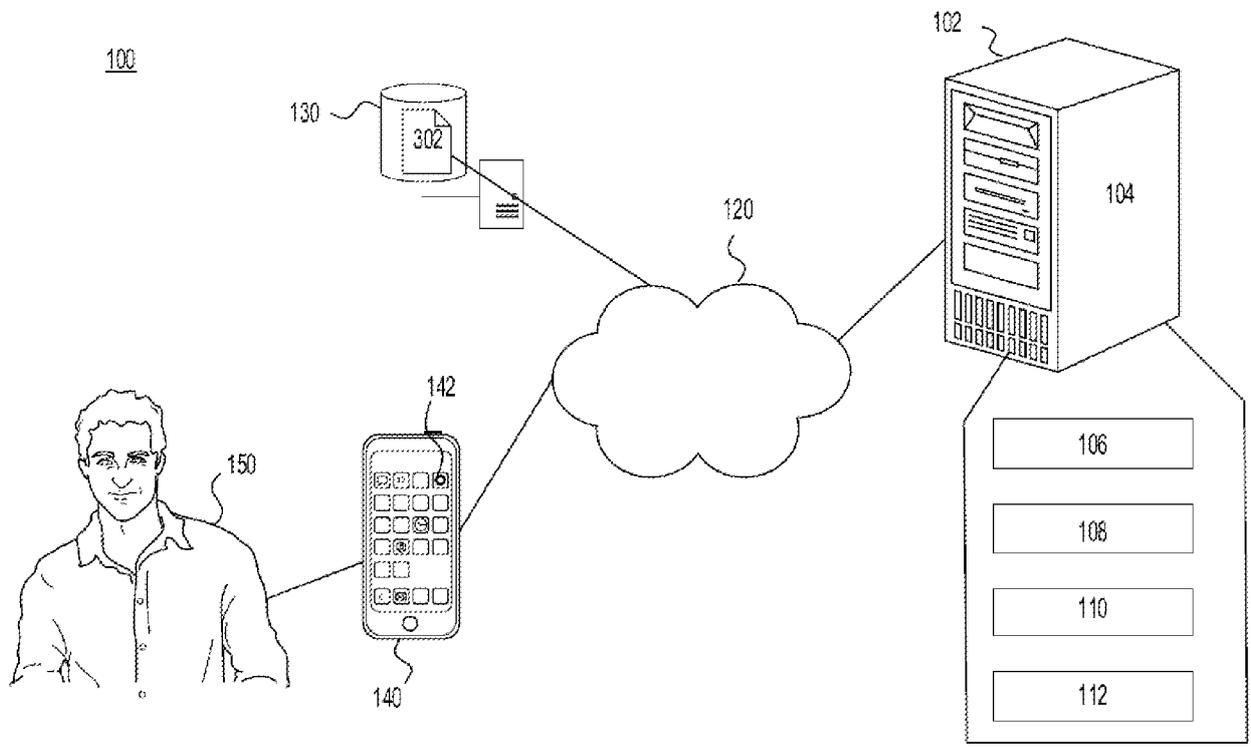
FIELD: information technology.

SUBSTANCE: in the method for creating a brief summary of digital content a reference to the digital content is received and syntactic analysis of content textual representation is carried out. The content is divided into an ordered set of fragments including the first and second fragments. A semantic analysis of each fragment is carried out and the utility parameter for each fragment and the relations between each pair of fragments are determined. In response to the fact that utility parameter of the second fragment exceeds the

predetermined threshold value, the second fragment is included in a subset of fragments for inclusion in the brief summary of digital content. In response to receiving a reference to the relation of the second fragment with the first one, the first fragment is included in a subset of fragments. A brief summary of digital content based on a subset of fragments is created.

EFFECT: expansion of the arsenal of technical means for creating brief summaries of digital content.

62 cl, 6 dwg



ФИГ. 1

RU 2637998 C1

RU 2637998 C1

[001] Настоящая технология относится к электронным устройствам и способам обработки цифрового контента (в частности, текстовых, аудио- и видеофайлов). Конкретнее, настоящее техническое решение относится к системам и способам для создания кратких изложений цифрового контента.

5 УРОВЕНЬ ТЕХНИКИ

[002] Разработки в сфере компьютерных технологий позволяют обрабатывать большие объемы данных, включая компьютерную обработку различных типов цифрового контента. Компьютерное создание кратких изложений цифрового контента является одной из сложных задач компьютерной лингвистики. Основная задача компьютерного создания кратких изложений цифрового контента обладает двумя аспектами: (i) скорость обработки (поскольку некоторые краткие изложения необходимо создавать «на лету») и (ii) точность (т.е. предоставление краткого изложения без потерь общего смысла).

[003] Из патентной заявки США №US 20140072285 (опубл. 13 марта 2014) известна система аннотирования цифрового контента. Анализатор взаимодействия зрителей принимает медиафайл, содержащий цифровой контент, медиафайл включает в себя множество сегментов. Сегмент медиафайла оценивается на основе взаимодействия множества оценщиков. Метрики просмотра на сегменте медиафайла измеряются на основе взаимодействия зрителей с сегментом медиафайла. Множество векторов признаков формируются на основе измеренных взаимодействий зрителей, где векторы признаков в наборе векторов признаков основаны на взаимодействии множества зрителей. Модель обучается на основе набора векторов признаков и оценки, присвоенной сегменту медиафайла. Модель применяется к сегментам медиафайла для формирования рейтинга интереса для сегментов медиафайла. Отредактированный медиафайл генерируется на основе сегментов медиафайла, имеющих рейтинги интереса, которые отвечают критерию.

[004] Из патентной заявки США №US2015027390 (опубл. 1 октября 2015) известна система и способ создания сжатой версии элемента мультимедиа контента. Мультимедийный элемент контента сжимается на основе его звуковой дорожки и желаемой степени сжатия. Звуковая дорожка извлекается и обрабатывается с помощью автоматического распознавания речи для получения упорядоченного по времени текстового представления. Текстовое представление разбивается на множество последовательностей сегментов. Оценка информативности, основанная на оценке характерных черт и оценке различий, вычисляется для каждого из сегментов. Оценка связанности также вычисляется для сегментов во множестве последовательностей. Подпоследовательность одной из последовательностей сегментов оптимизируется для информативности и согласованности и выбирается для создания нового элемента контента - сжатой версии мультимедийного элемента контента.

[005] Из патентной заявки США №US20140156651 (опубл. 5 июня 2014) известен способ автоматического создания сжатой версии медиаконтента. Указанный способ включает: получение от компьютерного устройства запроса на создание сжатой версии медиаконтента; сегментирование медиаконтента на множество сегментов; в котором по меньшей мере один из множества сегментов связан с метаданными; в котором метаданные содержат характеристики ранжирования релевантности каждого из множества сегментов; в котором ранжирование содержит оценки связанности по сегментам в соответствии с метаданными; и выбор по меньшей мере одного выбранного сегмента из указанного множества сегментов; в котором по меньшей мере один выбранный сегмент связан с наивысшей оценкой.

РАСКРЫТИЕ

[006] Задачей предлагаемой технологии является расширение арсенала средств указанного назначения и предложить новые способы и системы для создания кратких изложений цифрового контента. Также технология в некоторых своих вариантах реализации позволяет устранить ряд недостатков, которые присущи известному уровню техники.

[007] Первым объектом настоящей технологии является способ создания краткого изложения цифрового контента, осуществляемый процессором компьютерного устройства. Способ включает в себя: получение указания на цифровой контент; синтаксический анализ текстового представления цифрового контента; разделение цифрового контента на упорядоченное множество фрагментов, включающее в себя первый фрагмент и второй фрагмент; семантический анализ каждого фрагмента текстового представления цифрового контента; определение параметра полезности для каждого фрагмента из множества фрагментов; определение связи между каждой парой фрагментов из множества фрагментов; в ответ на то, что параметр полезности второго фрагмента превышает предварительно определенное пороговое значение параметра полезности, включение второго фрагмента в подмножество фрагментов для включения в краткое изложение цифрового контента; в ответ на получение указания на связь второго фрагмента с первым фрагментом включение первого фрагмента в подмножество фрагментов; создание краткого изложения цифрового контента на основе подмножества фрагментов.

[008] Возможен вариант осуществления способа, в котором разделение цифрового контента на упорядоченное множество фрагментов осуществляют по меньшей мере частично на основе предварительно обученного первого алгоритма машинного обучения.

[009] Возможен вариант осуществления способа, в котором первый алгоритм машинного обучения предварительно обучен разделять цифровой контент на множество фрагментов, причем в качестве входных параметров первого алгоритма машинного обучения задают данные синтаксического анализа текстового представления цифрового контента, а обучение первого алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

[0010] Возможен вариант осуществления способа, в котором параметр полезности каждого фрагмента из множества фрагментов определяют по меньшей мере частично на основе предварительно обученного второго алгоритма машинного обучения.

[0011] Возможен вариант осуществления способа, в котором второй алгоритм машинного обучения предварительно обучен определять параметр полезности каждого фрагмента из множества фрагментов, причем в качестве входных параметров второго алгоритма машинного обучения задают данные семантического анализа текстового представления цифрового контента, а обучение второго алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

[0012] Возможен вариант осуществления способа, в котором на этапе определения параметра полезности полезность по меньшей мере одного фрагмента из множества фрагментов оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра второго алгоритма машинного обучения.

[0013] Возможен вариант осуществления способа, в котором размер первого фрагмента не равен размеру второго фрагмента.

[0014] Возможен вариант осуществления способа, в котором в ответ на получение

указания на взаимосвязь второго фрагмента с первым фрагментом выполняют объединение первого и второго фрагмента в единый фрагмент.

5 [0015] Возможен вариант осуществления способа, в котором упорядоченное множество фрагментов включает в себя по меньшей мере три фрагмента: первый, второй и третий.

[0016] Возможен вариант осуществления способа, в котором указание на связь между каждыми двумя фрагментами из множества фрагментов представляет собой: параметр, указывающий на наличие связи между каждыми двумя фрагментами, либо параметр, указывающий на степень связанности между каждыми двумя фрагментами.

10 [0017] Возможен вариант осуществления способа, в котором определяют связь между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно обученного третьего алгоритма машинного обучения.

[0018] Возможен вариант осуществления способа, в котором третий алгоритм машинного обучения предварительно обучен определять связь между каждой парой 15 фрагментов из множества фрагментов, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные семантического и синтаксического анализов текстового представления цифрового контента, а обучение третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

20 [0019] Возможен вариант осуществления способа, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве 25 дополнительного входного параметра третьего алгоритма машинного обучения.

[0020] Возможен вариант осуществления способа, в котором определение связи между каждой парой фрагментов выполняют параллельно с этапом разделения цифрового контента на упорядоченное множество фрагментов.

[0021] Возможен вариант осуществления способа, в котором перед созданием краткого изложения цифрового контента на основе подмножества фрагментов 30 дополнительно осуществляют проверку связи между каждой парой фрагментов из подмножества фрагментов на основе предварительно обученного третьего алгоритма машинного обучения и/или на основе проверки по меньшей мере одним ассессором.

[0022] Возможен вариант осуществления способа, в котором указание на связь представляет собой параметр, указывающий степень связанности между фрагментами; 35 причем перед проверкой связи между каждой парой фрагментов из подмножества фрагментов задают пороговое значение параметра, указывающего степень связанности между фрагментами.

[0023] Возможен вариант осуществления способа, в котором параметр полезности первого фрагмента не превышает предварительно определенное пороговое значение.

40 [0024] Возможен вариант осуществления способа, в котором перед определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение параметра полезности; создают по меньшей мере одно дополнительное краткое изложение цифрового контента на основе по меньшей мере одного дополнительного подмножества фрагментов.

45 [0025] Возможен вариант осуществления способа, в котором по меньшей мере одно дополнительное краткое изложение цифрового контента отличается по размеру от краткого изложения цифрового контента.

[0026] Возможен вариант осуществления способа, в котором перед этапом разделения

цифрового контента получают запрос на создание краткого изложения цифрового контента.

5 [0027] Возможен вариант осуществления способа, в котором в ответ на получение запроса на создание цифрового контента отправляют ассессору по меньшей мере одно указание на фрагмент цифрового контента для оценки его параметра полезности; на этапе определения параметра полезности для каждого фрагмента из множества фрагментов в качестве дополнительного входного параметра используют упомянутую полученную оценку параметра полезности по меньшей мере одного из фрагментов.

10 [0028] Возможен вариант осуществления способа, в котором запрос на создание краткого изложения цифрового контента включает в себя пороговое значение параметра полезности фрагментов цифрового контента для включения в краткое изложение цифрового контента.

15 [0029] Возможен вариант осуществления способа, в котором запрос на создание краткого изложения цифрового контента включает в себя указание на степень владения тематикой цифрового контента.

[0030] Возможен вариант осуществления способа, в котором на этапе определения связи между фрагментами в качестве входного параметра используют полученное значение степени владения тематикой цифрового контента.

20 [0031] Возможен вариант осуществления способа, в котором цифровой контент представляет собой текстовый файл.

[0032] Возможен вариант осуществления способа, в котором цифровой контент представляет собой аудиофайл.

[0033] Возможен вариант осуществления способа, в котором после получения цифрового контента осуществляют создание текстового представления аудиофайла.

25 [0034] Возможен вариант осуществления способа, в котором при разделении аудиофайла на фрагменты и при создании указаний на взаимосвязь между фрагментами дополнительно выполняют анализ на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы.

30 [0035] Возможен вариант осуществления способа, в котором цифровой контент представляет собой видеофайл.

[0036] Возможен вариант осуществления способа, в котором после получения цифрового контента осуществляют создание текстового представления видеофайла.

35 [0037] Возможен вариант осуществления способа, в котором при разделении видеофайла на фрагменты и при определении взаимосвязи между фрагментами дополнительно выполняют анализ звукового представления на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы, а также анализ визуального представления на основе по меньшей мере одного из следующего: фон, сцена, цветность.

40 [0038] Другим объектом настоящей технологии является компьютерное устройство для создания краткого изложения цифрового контента, включающее в себя процессор, коммуникационный интерфейс, процессор функционально связан с модулем получения цифрового контента, модулем разделения цифрового контента на упорядоченное множество фрагментов, модулем определения параметра полезности, модулем определения связи между фрагментами, причем процессор выполнен с возможностью
45 осуществлять: получение указания на цифровой контент; синтаксический анализ текстового представления цифрового контента; разделение цифрового контента на упорядоченное множество фрагментов, включающее в себя первый фрагмент и второй фрагмент; семантический анализ каждого фрагмента текстового представления

цифрового контента; определение параметра полезности для каждого фрагмента из множества фрагментов; определение связи между каждой парой фрагментов из множества фрагментов; в ответ на то, что параметр полезности второго фрагмента превышает предварительно определенное пороговое значение параметра полезности, включение второго фрагмента в подмножество фрагментов для включения в краткое изложение цифрового контента; в ответ на получение указания на связь второго фрагмента с первым фрагментом включение первого фрагмента в подмножество фрагментов; создание краткого изложения цифрового контента на основе подмножества фрагментов.

10 [0039] Возможен вариант осуществления компьютерного устройства, в котором процессор выполнен с возможностью осуществлять разделение цифрового контента на упорядоченное множество фрагментов по меньшей мере частично на основе предварительно обученного первого алгоритма машинного обучения.

15 [0040] Возможен вариант осуществления компьютерного устройства, в котором первый алгоритм машинного обучения предварительно обучен разделять цифровой контент на множество фрагментов, причем в качестве входных параметров первого алгоритма машинного обучения задают данные синтаксического анализа текстового представления цифрового контента, а обучение первого алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

20 [0041] Возможен вариант осуществления компьютерного устройства, в котором процессор выполнен с возможностью определять параметр полезности каждого фрагмента из множества фрагментов по меньшей мере частично на основе предварительно обученного второго алгоритма машинного обучения.

25 [0042] Возможен вариант осуществления компьютерного устройства, в котором второй алгоритм машинного обучения предварительно обучен определять параметр полезности каждого фрагмента из множества фрагментов, причем в качестве входных параметров второго алгоритма машинного обучения задают данные семантического анализа текстового представления цифрового контента, а обучение второго алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей
30 выборки.

[0043] Возможен вариант осуществления компьютерного устройства, в котором на этапе определения параметра полезности полезность по меньшей мере одного фрагмента из множества фрагментов оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра второго алгоритма
35 машинного обучения.

[0044] Возможен вариант осуществления компьютерного устройства, в котором размер первого фрагмента не равен размеру второго фрагмента.

40 [0045] Возможен вариант осуществления компьютерного устройства, в котором в ответ на получение указания на взаимосвязь второго фрагмента с первым фрагментом процессор выполнен с возможностью выполнять объединение первого и второго фрагмента в единый фрагмент.

[0046] Возможен вариант осуществления компьютерного устройства, в котором упорядоченное множество фрагментов включает в себя по меньшей мере три фрагмента: первый, второй и третий.

45 [0047] Возможен вариант осуществления компьютерного устройства, в котором указание на связь между каждыми двумя фрагментами из множества фрагментов представляет собой: параметр, указывающий на наличие связи между каждыми двумя фрагментами, либо параметр, указывающий на степень связанности между каждыми

двумя фрагментами.

5 [0048] Возможен вариант осуществления компьютерного устройства, в котором процессор выполнен с возможностью определять связь между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно

10 обученного третьего алгоритма машинного обучения.
[0049] Возможен вариант осуществления компьютерного устройства, в котором третий алгоритм машинного обучения предварительно обучен определять связь между каждой парой фрагментов из множества фрагментов, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные семантического и синтаксического анализов текстового представления цифрового контента, а обучение

15 третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.
[0050] Возможен вариант осуществления компьютерного устройства, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра третьего алгоритма машинного обучения.

20 [0051] Возможен вариант осуществления компьютерного устройства, в котором процессор выполнен с возможностью определения связи между каждой парой фрагментов параллельно с этапом разделения цифрового контента на упорядоченное множество фрагментов.

25 [0052] Возможен вариант осуществления компьютерного устройства, в котором перед созданием краткого изложения цифрового контента на основе подмножества фрагментов процессор выполнен с возможностью дополнительно осуществлять проверку связи между каждой парой фрагментов из подмножества фрагментов на основе предварительно обученного третьего алгоритма машинного обучения и/или на основе проверки по меньшей мере одним ассессором.

30 [0053] Возможен вариант осуществления компьютерного устройства, в котором указание на связь представляет собой параметр, указывающий степень связанности между фрагментами; причем перед проверкой связи между каждой парой фрагментов из подмножества фрагментов задают пороговое значение параметра, указывающего степень связанности между фрагментами.

35 [0054] Возможен вариант осуществления компьютерного устройства, в котором параметр полезности первого фрагмента не превышает предварительно определенное пороговое значение.

40 [0055] Возможен вариант осуществления компьютерного устройства, в котором перед определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение параметра полезности; процессор выполнен с возможностью создавать по меньшей мере одно дополнительное краткое изложение цифрового контента на основе по меньшей мере одного дополнительного подмножества фрагментов.

45 [0056] Возможен вариант осуществления компьютерного устройства, в котором по меньшей мере одно дополнительное краткое изложение цифрового контента отличается по размеру от краткого изложения цифрового контента.

[0057] Возможен вариант осуществления компьютерного устройства, в котором процессор выполнен с возможностью перед этапом разделения цифрового контента получать запрос на создание краткого изложения цифрового контента.

[0058] Возможен вариант осуществления компьютерного устройства, в котором в ответ на получение запроса на создание цифрового контента процессор выполнен с возможностью отправлять ассессору по меньшей мере одно указание на фрагмент цифрового контента для оценки его параметра полезности; на этапе определения параметра полезности для каждого фрагмента из множества фрагментов в качестве дополнительного входного параметра используют упомянутую полученную оценку параметра полезности по меньшей мере одного из фрагментов.

[0059] Возможен вариант осуществления компьютерного устройства, в котором запрос на создание краткого изложения цифрового контента включает в себя пороговое значение параметра полезности фрагментов цифрового контента для включения в краткое изложение цифрового контента.

[0060] Возможен вариант осуществления компьютерного устройства, в котором запрос на создание краткого изложения цифрового контента включает в себя указание на степень владения тематикой цифрового контента.

[0061] Возможен вариант осуществления компьютерного устройства, в котором на этапе определения связи между фрагментами в качестве входного параметра используют полученное значение степени владения тематикой цифрового контента.

[0062] Возможен вариант осуществления компьютерного устройства, в котором цифровой контент представляет собой текстовый файл.

[0063] Возможен вариант осуществления компьютерного устройства, в котором цифровой контент представляет собой аудиофайл.

[0064] Возможен вариант осуществления компьютерного устройства, в котором после получения цифрового контента процессор выполнен с возможностью осуществлять создание текстового представления аудиофайла.

[0065] Возможен вариант осуществления компьютерного устройства, в котором при разделении аудиофайла на фрагменты и при создании указаний на взаимосвязь между фрагментами процессор выполнен с возможностью дополнительно выполнять анализ на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы.

[0066] Возможен вариант осуществления компьютерного устройства, в котором цифровой контент представляет собой видеофайл.

[0067] Возможен вариант осуществления компьютерного устройства, в котором после получения цифрового контента процессор выполнен с возможностью осуществлять создание текстового представления видеофайла.

[0068] Возможен вариант осуществления компьютерного устройства, в котором при разделении видеофайла на фрагменты и при определении взаимосвязи между фрагментами процессор выполнен с возможностью дополнительно выполнять анализ звукового представления на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы, а также анализ визуального представления на основе по меньшей мере одного из следующего: фон, сцена, цветность.

[0069] Настоящая технология позволяет создавать краткие содержания (сокращенные версии) цифрового контента с максимальной полезностью для пользователя. Причем краткое содержание цифрового контента предварительно не ограничивается по длительности, что позволяет использовать любое необходимое количество фрагментов для получения полезного и связного краткого изложения цифрового контента.

[0070] В контексте настоящего описания, если четко не указано иное, "электронное устройство", "компьютерное устройство", "клиентское устройство", "сервер", "удаленный

сервер" и "компьютерная система" подразумевают под собой аппаратное и/или системное обеспечение, подходящее к решению соответствующей задачи. Таким образом, некоторые неограничивающие примеры аппаратного и/или программного обеспечения включают в себя компьютеры (серверы, настольные компьютеры, ноутбуки, нетбуки и так далее), смартфоны, планшеты, сетевое оборудование (маршрутизаторы, коммутаторы, шлюзы и так далее) и/или их комбинацию.

[0071] В контексте настоящего описания, если четко не указано иное, «машиночитаемый носитель», «хранилище», «устройство хранения» подразумевает под собой носитель абсолютно любого типа и характера, не ограничивающие примеры включают в себя ОЗУ, ПЗУ, диски (компакт диски, DVD-диски, дискеты, жесткие диски и т.д.), USB-ключи, флеш-карты, твердотельные накопители и накопители на магнитной ленте.

[0072] В контексте настоящего описания, если четко не указано иное, термин «модуль» подразумевает под собой программное обеспечение (соответствующее конкретному аппаратному контексту), которое является необходимым и достаточным для выполнения конкретной(ых) указанной(ых) функции(й).

[0073] В контексте настоящего описания, если четко не указано иное, «указание» информационного элемента может представлять собой сам информационный элемент или указатель, отсылку, ссылку или другой косвенный способ, позволяющий получателю указания найти сеть, память, базу данных или другой машиночитаемый носитель, из которого может быть извлечен информационный элемент. Например, указание на документ может включать в себя сам документ (т.е. его содержимое), или же оно может являться уникальным дескриптором документа, идентифицирующим файл по отношению к конкретной файловой системе, или каким-то другими средствами передавать получателю указание на сетевую папку, адрес памяти, таблицу в базе данных или другое место, в котором можно получить доступ к файлу. Как будет понятно специалистам в данной области техники, степень точности, необходимая для такого указания, зависит от степени первичного понимания того, как должна быть интерпретирована информация, которой обмениваются получатель и отправитель указателя. Например, если до установления связи между отправителем и получателем понятно, что указание информационного элемента принимает вид ключа базы данных для записи в конкретной таблице заранее установленной базы данных, включающей в себя информационный элемент, то передача ключа базы данных – это все, что необходимо для эффективной передачи информационного элемента получателю, несмотря на то, что сам по себе информационный элемент не передавался между отправителем и получателем указания.

[0074] В контексте настоящего описания, если конкретно не указано иное, слова «первый», «второй», «третий» и т.д. используются в виде прилагательных исключительно для того, чтобы отличать существительные, к которым они относятся, друг от друга, а не для целей описания какой-либо конкретной связи между этими существительными. Так, например, следует иметь в виду, что использование терминов «первый сервер» и «третий сервер» не подразумевает какого-либо порядка, отнесения к определенному типу, хронологии, иерархии или ранжирования (например) серверов/между серверами, равно как и их использование (само по себе) не предполагает, что некий «второй сервер» обязательно должен существовать в той или иной ситуации. В дальнейшем, как указано здесь в других контекстах, упоминание «первого» элемента и «второго» элемента не исключает возможности того, что это один и тот же фактический реальный элемент. Так, например, в некоторых случаях, «первый» сервер и «второй» сервер могут являться одним и тем же программным и/или аппаратным обеспечением, а в других случаях они

могут являться разным программным и/или аппаратным обеспечением.

[0075] Под термином «параметр полезности фрагмента цифрового контента» в контексте настоящей технологии понимается выражение степени интересности, важности, информативности фрагмента цифрового контента для пользователя.

5 Например, в видеолекции фрагмент с введением и фрагмент с выводами обычно будут обладать большей полезностью для пользователя, чем фрагменты, на которых лектор делает отступления от основной темы для привлечения внимания аудитории (например, рассказывает анекдот).

[0076] Под термином «взаимосвязь фрагментов», «связанность фрагментов», «связь
10 фрагментов» в контексте настоящей технологии понимается понятность для пользователя одного фрагмента при исключении другого фрагмента. Например, в первом фрагменте цифрового контента дано описание учебной задачи, а во втором фрагменте дан ответ на указанную учебную задачу. В таком случае второй фрагмент будет непонятен без первого, то есть связан с ним. При этом первый фрагмент может
15 быть понятен без второго фрагмента.

[0077] Дополнительные и/или альтернативные характеристики, аспекты и преимущества вариантов осуществления настоящей технологии станут очевидными из последующего описания, прилагаемых чертежей и прилагаемой формулы изобретения.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0078] Для лучшего понимания настоящей технологии, а также других ее аспектов и характерных черт сделана ссылка на следующее описание, которое должно использоваться в сочетании с прилагаемыми чертежами, где:

[0079] на Фиг. 1 представлена принципиальная схема системы, выполненной в соответствии с вариантами осуществления настоящей технологии, не ограничивающими
25 ее объем;

[0080] на Фиг. 2 представлена блок-схема способа 200, выполняемого в рамках системы, изображенной на Фиг. 1, и выполненного в соответствии с вариантами осуществления настоящей технологии, не ограничивающими ее объем;

[0081] на Фиг. 3 схематично показан этап разделения цифрового контента на
30 упорядоченное множество фрагментов в соответствии с неограничивающими вариантами осуществления настоящей технологии;

[0082] на Фиг. 4 схематично показан этап определения параметра полезности определения параметра полезности для каждого фрагмента из множества фрагментов в соответствии с неограничивающими вариантами осуществления настоящей технологии;

[0083] на Фиг. 5 схематично показан этап определение связи между каждой парой
35 фрагментов из множества фрагментов в соответствии с неограничивающими вариантами осуществления настоящей технологии;

[0084] на Фиг. 6 схематично показан этап создания краткого изложения цифрового контента на основе подмножества фрагментов в соответствии с неограничивающими
40 вариантами осуществления настоящей технологии.

ОСУЩЕСТВЛЕНИЕ

[0085] На Фиг. 1 представлена принципиальная схема системы 100, выполненной в соответствии с вариантами осуществления настоящей технологии, не ограничивающими ее объем. Важно иметь в виду, что нижеследующее описание системы 100 представляет собой описание иллюстративных вариантов осуществления настоящей технологии.
45 Таким образом, все последующее описание представлено только как описание иллюстративного примера настоящей технологии. Это описание не предназначено для определения объема или установления границ настоящей технологии. Некоторые

полезные примеры модификаций системы 100 также могут быть охвачены
нижеследующим описанием. Целью этого является также исключительно помощь в
понимании, а не определение объема и границ настоящей технологии. Эти модификации
не представляют собой исчерпывающий список, и специалистам в данной области
5 техники будет понятно, что возможны и другие модификации. Кроме того, это не
должно интерпретироваться так, что там, где не были изложены примеры модификаций,
никакие модификации невозможны, и/или что то, что описано, является единственным
вариантом осуществления этого экземпляра настоящей технологии. Как будет понятно
10 специалисту в данной области техники, это, скорее всего, не так. Кроме того, следует
иметь в виду, что система 100 представляет собой в некоторых конкретных проявлениях
достаточно простой вариант осуществления настоящей технологии, и в подобных
случаях представлен здесь с целью облегчения понимания. Как будет понятно
специалисту в данной области техники, многие варианты осуществления настоящей
технологии будут обладать гораздо большей сложностью.

15 [0086] Со ссылкой на Фиг. 1 система 100 включает в себя компьютерное устройство
102, связанное по сети передачи данных 120 с по меньшей мере одним сервером
источника цифрового контента 130 и по меньшей мере одним клиентским устройством
140, ассоциированным с пользователем 150. Компьютерное устройство 102 включает
в себя процессор 104, коммуникационный интерфейс (не показан), процессор 104
20 функционально связан с модулем 106 получения цифрового контента, модулем 108
разделения цифрового контента на упорядоченное множество фрагментов, модулем
110 определения параметра полезности и модулем 112 определения связи между
фрагментами. В некоторых вариантах осуществления настоящей технологии, не
ограничивающих ее объем, сеть передачи данных 120 может представлять собой
25 Интернет. В других вариантах осуществления настоящей технологии, сеть передачи
данных 120 может быть реализована иначе - в виде глобальной сети связи, локальной
сети связи, частной сети связи и т.п.

[0087] Согласно неограничивающему варианту осуществления системы 100 на Фиг.
1 модули 106, 108, 110 и 112 являются частью компьютерного устройства 102. Однако
30 возможны варианты осуществления системы 100, в которых по меньшей мере один из
модулей 106, 108, 110, 112 не является частью компьютерного устройства 102, а,
например, находится на отдельном сервере (не показан) и связан с компьютерным
устройством 102 по сети передачи данных 120. Таким образом, то, как именно процессор
104 компьютерного устройства 102 функционально связан с модулями 106, 108, 110 и
35 112, не является ограничивающим условием настоящей технологии. При использовании
модулей 106, 108, 110, и 112 процессор 104 выполнен с возможностью осуществлять
различные операции, описанные ниже.

[0088] Компьютерное устройство 102 может представлять собой обычный
компьютерный сервер, например сервер Dell™ PowerEdge™, на котором используется
40 операционная система Microsoft™ Windows Server™. Важно иметь в виду, что
компьютерное устройство 102 может быть реализовано на любом подходящем
аппаратном и/или прикладном программном, и/или системном программном
обеспечении или их комбинации. В представленном варианте осуществления настоящей
технологии, не ограничивающем ее объем, компьютерное устройство 102 является
45 одиночным сервером. В других вариантах осуществления настоящей технологии, не
ограничивающих ее объем, функциональность компьютерного устройства 102 обработки
может быть разделена и может выполняться с помощью нескольких компьютерных
устройств.

[0089] Модификации коммуникационного интерфейса (не показан) хорошо известны в данной области техники. То, как именно компьютерное устройство 102 соединено с сервером источника цифрового контента 130 и клиентским устройством 140 никак конкретно не ограничено. В качестве примера, но не ограничения, компьютерное устройство 102 может быть связано с сервером источника цифрового контента 130 по проводной сети передачи данных (соединение на основе сети Ethernet). Компьютерное устройство 102 может быть связано с по меньшей мере одним клиентским устройством 140 через беспроводную сеть связи (например, среди прочего, соединение сети 3G, соединение сети 4G, беспроводной интернет Wireless Fidelity или коротко WiFi®, Bluetooth® и тому подобное). Альтернативно или дополнительно компьютерное устройство 102 может быть соединено с сервером источника цифрового контента 130, клиентским устройством 140 с использованием проводной или беспроводной сети передачи данных либо их комбинации.

[0090] Сервер источника цифрового контента 130 может представлять собой обычный компьютерный сервер, например сервер Dell™ PowerEdge™, на котором используется операционная система Microsoft™ Windows Server™. Важно иметь в виду, что сервер источника цифрового контента 130 может быть реализован на любом подходящем аппаратном, и/или прикладном программном, и/или системном программном обеспечении или их комбинации. В представленном варианте осуществления настоящей технологии, не ограничивающем ее объем, сервер источника цифрового контента 130 является одиночным сервером. В других вариантах осуществления настоящей технологии, не ограничивающих ее объем, функциональность сервера источника цифрового контента 130 может быть разделена и может выполняться с помощью нескольких серверов. Важно иметь в виду, что сервер источника цифрового контента 130 и компьютерное устройство 102 не обязательно должны быть реализованы одинаковым образом.

[0091] Сервер источника цифрового контента 130 размещает содержимое, включая один или несколько цифровых документов 302. Природа одного или нескольких цифровых документов 302 никак конкретно не ограничена. Так в качестве цифрового документа 302 может быть использован различный цифровой контент, в частности: текстовый файл, аудиофайл, видеофайл. Цифровой документ 302 может представлять собой, например, видеолекцию «Analysis of Continuous Systems» by Professor K. J. Bathe (Анализ непрерывных систем). Еще одним неограничивающим примером цифрового документа 302 может быть веб-страница, содержащая текст (например, статья из WIKIPEDIA™ о датацентрах). Цифровой документ 302 может представлять собой, в качестве другого примера, аудиоверсию книги. Например, аудиоверсию книги «Гордость и Предубеждение» Джейн Остин.

[0092] Возможен вариант осуществления настоящей технологии, в котором компьютерное устройство 102 дополнительно включает в себя хранилище данных (не показано), на котором расположен цифровой контент (в частности, цифровой документ 302). Таким образом сервер источника цифрового контента 130 и компьютерное устройство 102 в некоторых неограничивающих вариантах осуществления настоящей технологии могут представлять собой единое устройство.

[0093] Вне зависимости от типа и содержания цифрового документа 302, может быть желательно создать краткое изложение цифрового документа 302. Вариантов использования такого краткого изложения бесчисленное множество. Например, может быть желательно создать краткое изложение цифрового документа 302, который представляет собой веб-ресурс для включения такого созданного краткого изложения

на страницу результатов поиска (SERP), которая была создана сервером поисковой системы в ответ на поисковый запрос пользователя. В качестве другого примера, может быть желательно создать краткое изложение цифрового документа 302, который является видео- или аудиофайлом для снижения трафика сети передачи данных, при сохранении качества контента и его основного смысла. При этом создание кратких изложений цифрового контента позволяет увеличить и скорость ознакомления пользователя с цифровым контентом, а следовательно, существенно снижать трафик сети передачи данных и нагрузку на сервер источника цифрового контента 130.

[0094] Вне зависимости от того, по какой причине может быть желательно подобное обобщение, в системе 100 предусмотрено клиентское устройство 140. Клиентское устройство 140 способно получать доступ к серверу источника цифрового контента 130 и/или компьютерному устройству 102 через сеть 120 передачи данных. Клиентское устройство 140 способно получать доступ к серверу источника цифрового контента 130 и/или компьютерному устройству 102 либо для доступа к содержимому одного или нескольких цифровых документов 302, либо для доступа к их созданным кратким изложениям. С этой целью, клиентское устройство 140 соединено с сетью 120 через линию передачи данных (не пронумерована).

[0095] Аппаратное осуществление клиентского устройства 140 никак конкретно не ограничено. Клиентское устройство 140 может, например, в качестве не ограничивающего примера, представлять собой настольный компьютер, ноутбук, смартфон, как показано на Фиг. 1 (например, Apple iPhoneTM или Samsung Galaxy S5TM), персональный цифровой органайзер (PDA) или другое устройство, включающее в себя функциональность вычисления и возможности передачи данных. Клиентское устройство 140 может включать в себя, среди прочего, внутренние аппаратные компоненты, включая один или несколько одно- или мультитядерных процессоров, которые коллективно представлены здесь как процессор (не показан), ОЗУ (не показано), сетевой интерфейс (не показано) для связи с сервером источника цифрового контента 130.

[0096] Реализация линии передачи данных не ограничена и будет зависеть от того, какое клиентское устройство 140 используется. В качестве примера (но не ограничения) в данных вариантах осуществления настоящего технического решения в случаях, когда клиентское устройство 140 представляет собой беспроводное устройство связи (например, смартфон), линия передачи данных представляет собой беспроводную сеть передачи данных (например, среди прочего, линию передачи данных 3G, линию передачи данных 4G, беспроводной интернет Wireless Fidelity или WiFi®, Bluetooth® и т.п.). В тех примерах, где клиентское устройство 140 представляет собой портативный компьютер, линия передачи данных может быть как беспроводной (беспроводной интернет Wireless Fidelity или WiFi®, Bluetooth® и т.п.), так и проводной (соединение на основе сети Ethernet).

[0097] Клиентское устройство 140 связано с пользователем 150, который может взаимодействовать с клиентским устройством 140. В некоторых вариантах осуществления настоящего технического решения, клиентское устройство 140 может быть выполнено с возможностью выполнять браузерное приложение (не показано). В общем случае, задачей браузерного приложения является предоставление пользователю 150 доступа к одному или нескольким веб-ресурсам и/или к одному или нескольким сервисам. Один или несколько веб-ресурсов и/или один или несколько сервисов могут храниться или предоставляться компьютерным устройством 102 обработки и/или сервером источника цифрового контента 130.

[0098] Реализация браузерного приложения никак конкретно не ограничена.

Например, браузерное приложение 142 может быть реализовано как Яндекс™ браузер. Несмотря на то, что клиентское устройство 140 связано с пользователем 150, эта связь не обязательно требует от пользователя 150 авторизации или чего-либо подобного.

5 [0099] Важно иметь в виду, что варианты реализации клиентского устройства 140, линии передачи данных и сети 120 передачи данных приведены исключительно для наглядности. Таким образом, специалисты в данной области техники смогут понять подробности других конкретных вариантов осуществления клиентского устройства 140, линии передачи данных и сети 120 передачи данных. То есть представленные здесь
10 примеры не ограничивают объем настоящей технологии. Возможен вариант осуществления настоящей технологии, в котором компьютерное устройство 102 выполнено с возможностью получения запроса от клиентского устройства 140 на создание краткого изложения цифрового контента.

[00100] Со ссылкой на Фиг. 1 компьютерное устройство 102 включает в себя модуль получения цифрового контента 106. Модуль получения цифрового контента 106 может
15 быть в данном контексте физическим устройством, аппаратом или множеством модулей, выполненных с использованием аппаратного обеспечения, например с помощью интегральной схемы специального назначения (ASIC), или программируемой логической интегральной схемы (FPGA), или же комбинации аппаратного и программного
20 обеспечения, например с помощью микропроцессорной системы и набора инструкций, реализующих функционал модуля, которые (при выполнении) трансформируют микропроцессорную систему в устройство специального назначения. Модуль получения цифрового контента 106 также может быть реализован в виде комбинации аппаратного и программного обеспечения, причем некоторые конкретные функции реализуются за
25 счет аппаратного обеспечения, а другие функции реализуются комбинацией аппаратного и программного обеспечения.

[00101] В других вариантах осуществления настоящей технологии модуль получения цифрового контента 106 может также быть реализован как программное обеспечение для выполнения различными типами процессоров. Модуль получения цифрового
30 контента 106 может, например, включать один или несколько физических или логических блоков компьютерных инструкций, которые могут быть организованы, например, как объект, процедура или функция. Однако данные исполнимые модули не должны быть физически расположены вместе и они могут включать в себя различные инструкции, сохраненные в различных местах, которые, при логическом соединении друг с другом,
35 включают в себя данный модуль и выполняют заявленную задачу для данного модуля.

[00102] Модуль получения цифрового контента 106 может получать указание на цифровой контент и/или сам цифровой контент (в частности, цифровой документ 302 и/или указание на него) от сервера источника цифрового контента 130 по сети передачи
40 данных 120 и/или от клиентского устройства 140.

[00103] В некоторых вариантах осуществления настоящей технологии, когда цифровой документ 302 представляет собой не текстовый формат, модуль получения цифрового
45 контента 106 может дополнительно получать текстовое представление цифрового контента (цифрового документа 302). Так, например, цифровой документ 302, представляющий собой видеолекцию, может дополнительно содержать текстовое представление (стенограмму) цифрового документа 302, хранящуюся на сервере источника цифрового контента 130. Модуль получения цифрового контента 106 может
дополнительно быть выполнен с возможностью создания текстового представления цифрового документа 302, представляющего собой аудиофайл, видеофайл или изображение. Таким образом, модуль получения цифрового контента 106 может

включать в себя машиночитаемые инструкции, при выполнении которых процессор 104 компьютерного устройства 102 способен выполнять распознавание речи и/или изображений для создания текстового представления цифрового контента, в частности цифрового документа 302. Алгоритмы распознавания речи широко известны в данной области техники, однако в качестве примера такой алгоритм может быть реализован посредством глубокой нейронной сети, которая позволяет учитывать множество взаимозависимых речевых атрибутов. Реализация глубокой нейронной сети никак конкретно не ограничена.

[00104] Со ссылкой на Фиг. 1 компьютерное устройство 102 включает в себя модуль разделения цифрового контента 108. Модуль разделения цифрового контента 108 может быть в данном контексте физическим устройством, аппаратом или множеством модулей, выполненных с использованием аппаратного обеспечения, например с помощью интегральной схемы специального назначения (ASIC), или программируемой логической интегральной схемы (FPGA), или же комбинации аппаратного и программного обеспечения, например с помощью микропроцессорной системы и набора инструкций, реализующих функционал модуля, которые (при выполнении) трансформируют микропроцессорную систему в устройство специального назначения. Модуль разделения цифрового контента 108 также может быть реализован в виде комбинации аппаратного и программного обеспечения, причем некоторые конкретные функции реализуются за счет аппаратного обеспечения, а другие функции реализуются комбинацией аппаратного и программного обеспечения.

[00105] В других вариантах осуществления настоящей технологии модуль разделения цифрового контента 108 может также быть реализован как программное обеспечение для выполнения различными типами процессоров. Модуль разделения цифрового контента 108 может, например, включать один или несколько физических или логических блоков компьютерных инструкций, которые могут быть организованы, например, как объект, процедура или функция. Однако данные исполнимые модули не должны быть физически расположены вместе и они могут включать в себя различные инструкции, сохраненные в различных местах, которые, при логическом соединении друг с другом, включают в себя данный модуль и выполняют заявленную задачу для данного модуля.

[00106] Со ссылкой на Фиг. 1 и Фиг. 3 модуль разделения цифрового контента 108 выполнен с возможностью осуществлять синтаксический анализ (парсинг) текстового представления цифрового контента (в частности, цифрового документа 302). По меньшей мере на основе синтаксического анализа текстового представления цифрового контента модуль разделения цифрового контента 108 выполнен с возможностью осуществлять разделение цифрового контента (цифрового документа 302) на упорядоченное множество фрагментов, включающее в себя первый фрагмент 3021 и второй фрагмент 3022. Возможен вариант осуществления настоящей технологии, в котором размер первого фрагмента 3021 не равен размеру второго фрагмента 3022. В альтернативном варианте осуществления настоящей технологии размеры некоторых фрагментов, например первого фрагмента 3021 и второго фрагмента 3022, могут совпадать. Возможен вариант осуществления настоящей технологии, в котором упорядоченное множество фрагментов включает в себя по меньшей мере три фрагмента: первый 3021, второй 3022 и третий 3023. Как показано в иллюстративном примере на Фиг. 3 упорядоченное множество фрагментов может включать в себя и большее количество фрагментов, в частности 8 фрагментов (3021, 3022, 3023, 3024, 3025, 3026, 3027, 3028). Стоит отметить, что для целей настоящей технологии количество и размер фрагментов, на которые процессор 104 посредством модуля разделения цифрового контента 108

выполняет разделение цифрового контента, не является ограничивающим условием. Так, например, цифровой документ 302 может быть разделен на 8 фрагментов, как показано на Фиг.3, или, например, на 100 фрагментов (не показано), или всего на 3 фрагмента (не показано).

5 [00107] Кроме того, модуль разделения цифрового контента 108 может быть выполнен с возможностью анализа звукового представления цифрового контента и определять, в частности: интонацию, музыкальное сопровождение, тембр голоса, темп голоса, тон
10 голоса, паузы, и т.д. Модуль разделения цифрового контента 108 может быть так же выполнен с возможностью анализа визуального представления цифрового контента и определять, в частности: фон, сцену, цветность, распознавать образы и т.д.

[00108] В некоторых вариантах осуществления настоящей технологии модуль разделения цифрового контента 108 выполнен с возможностью осуществлять разделение цифрового контента на упорядоченное множество фрагментов по меньшей мере
15 частично на основе предварительно обученного первого алгоритма машинного обучения. В некоторых вариантах осуществления настоящей технологии первый алгоритм машинного обучения предварительно обучен разделять контент на множество фрагментов, причем в качестве входных параметров первого алгоритма машинного обучения задают данные парсинга (синтаксического анализа) текстового представления
20 цифрового контента, а обучение первого алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки. В качестве дополнительных входных параметров первого алгоритма машинного обучения могут быть заданы результаты анализа звукового и/или визуального представления цифрового контента.

[00109] В некоторых вариантах осуществления настоящей технологии, в которых цифровой документ 302 представляет собой аудио- или видеофайл, модуль разделения
25 цифрового контента 108 выполнен с возможностью выполнять дополнительный анализ звукового представления цифрового контента на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы. Причем первый алгоритм машинного обучения дополнительно
30 предварительно обучают разделять цифровой контент на множество фрагментов на основе анализа звукового представления цифрового контента.

[00110] Например, для случая, когда цифровой документ 302 представляет собой аудиокнигу, определенную роль при разделении цифрового контента на упорядоченное
множество фрагментов может играть звуковое представление цифрового контента, в частности наличие пауз и их длительность между предложениями, изменение
35 характеристик голоса диктора и др.

[00111] В некоторых вариантах осуществления настоящей технологии, в которых цифровой документ 302 представляет собой видеофайл модуль разделения цифрового
контента 108 выполнен с возможностью осуществлять дополнительный анализ
40 визуального представления цифрового контента на основе по меньшей мере одного из следующего: фон, сцена, цветность и др. Причем первый алгоритм машинного обучения дополнительно предварительно обучают разделять цифровой контент на множество фрагментов на основе анализа визуального представления цифрового контента.

[00112] Например, для случая, когда цифровой документ 302 представляет собой видеолекцию, определенную роль при разделении цифрового контента на упорядоченное
45 множество фрагментов может играть визуальное представление цифрового контента, в частности сменяемость сцены (лектор у доски, лектор, аудитория, доска).

[00113] Со ссылкой на Фиг. 1 компьютерное устройство 102 включает в себя модуль определения параметра полезности 110. Модуль определения параметра полезности

110 может быть в данном контексте физическим устройством, аппаратом или множеством модулей, выполненных с использованием аппаратного обеспечения, например с помощью интегральной схемы специального назначения (ASIC), или программируемой логической интегральной схемы (FPGA), или же комбинации аппаратного и программного обеспечения, например с помощью микропроцессорной системы и набора инструкций, реализующих функционал модуля, которые (при выполнении) трансформируют микропроцессорную систему в устройство специального назначения. Модуль определения параметра полезности 110 также может быть реализован в виде комбинации аппаратного и программного обеспечения, причем некоторые конкретные функции реализуются за счет аппаратного обеспечения, а другие функции реализуются комбинацией аппаратного и программного обеспечения.

[00114] В других вариантах осуществления настоящей технологии модуль определения параметра полезности 110 может также быть реализован как программное обеспечение для выполнения различными типами процессоров. Модуль определения параметра полезности 110 может, например, включать один или несколько физических или логических блоков компьютерных инструкций, которые могут быть организованы, например, как объект, процедура или функция. Однако данные исполнимые модули не должны, но могут быть физически расположены вместе и они могут включать в себя различные инструкции, сохраненные в различных местах, которые, при логическом соединении друг с другом, включают в себя данный модуль и выполняют заявленную задачу для данного модуля.

[00115] Модуль определения параметра полезности 110 выполнен с возможностью осуществлять семантический анализ текстового представления цифрового контента (в частности, цифрового документа 302). По меньшей мере на основе семантического анализа текстового представления цифрового контента модуль определения параметра полезности 110 выполнен с возможностью осуществлять определение параметра полезности для каждого фрагмента из множества фрагментов.

[00116] В некоторых вариантах осуществления настоящей технологии модуль определения параметра полезности 110 выполнен с возможностью осуществлять определение параметра полезности для каждого фрагмента из множества фрагментов по меньшей мере частично на основе предварительно обученного второго алгоритма машинного обучения. В некоторых вариантах осуществления настоящей технологии второй алгоритм машинного обучения предварительно обучен выполнять определение параметра полезности для каждого фрагмента из множества фрагментов, причем в качестве входных параметров второго алгоритма машинного обучения задают данные семантического анализа текстового представления каждого фрагмента цифрового контента, а обучение второго алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

[00117] В некоторых вариантах осуществления настоящей технологии, в которых цифровой документ 302 представляет собой аудио- или видеофайл, модуль определения параметра полезности выполнен с возможностью выполнять дополнительный анализ звукового представления цифрового контента на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы. Причем второй алгоритм машинного обучения дополнительно предварительно обучают определять параметр полезности каждого фрагмента из множества фрагментов на основе анализа звукового представления цифрового контента.

[00118] Например, для случая, когда цифровой документ 302 представляет собой аудиокнигу, определенную роль при определении параметра полезности каждого

фрагмента из множества фрагментов цифрового контента может играть звуковое представление цифрового контента, в частности интонация диктора, повышение тона, наличие пауз и их длительность и др. Например, посредством паузы может быть выделено наиболее значимое слово в предложении.

5 [00119] Например, для случая, когда цифровой документ 302 представляет собой видеолекцию, определенную роль при определении параметра полезности каждого фрагмента из множества фрагментов может играть наличие повторов, усиление голоса лектора, паузы, замедление при диктовке наиболее важных фрагментов и т.д. Например, наиболее важные предложения лектор может читать медленнее.

10 [00120] В некоторых вариантах осуществления настоящей технологии, в которых цифровой документ 302 представляет собой видеофайл, модуль определения параметра полезности 110 выполнен с возможностью осуществлять дополнительный анализ визуального представления цифрового контента на основе по меньшей мере одного из следующего: фон, сцена, цветность, образы и др. Причем второй алгоритм машинного обучения дополнительно предварительно обучают определять параметр полезности фрагментов на основе анализа визуального представления цифрового контента.

15 [00121] Например, для случая, когда цифровой документ 302 представляет собой видеолекцию, определенную роль при определении параметра полезности каждого фрагмента из множества фрагментов может играть визуальное представление цифрового контента, в частности крупный план доски или слайда презентации.

[00122] Возможен вариант осуществления настоящей технологии, в котором на этапе определения параметра полезности полезность по меньшей мере одного фрагмента из множества фрагментов оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра второго алгоритма машинного обучения. Так процессор 104 посредством модуля определения параметра полезности 110 выполнен с возможностью отправки по меньшей мере одного фрагмента из множества фрагментов цифрового документа 302 человеку-ассессору для оценки полезности данного фрагмента. Оценка, полученная от ассессора, может быть использована в качестве дополнительного входного параметра второго алгоритма машинного обучения для определения полезности каждого фрагмента из множества фрагментов цифрового документа 302 и/или другого цифрового контента.

30 [00123] Со ссылкой на Фиг. 1 компьютерное устройство 102 включает в себя модуль определения связи между фрагментами 112. Модуль определения связи между фрагментами 112 может быть в данном контексте физическим устройством, аппаратом или множеством модулей, выполненных с использованием аппаратного обеспечения, например с помощью интегральной схемы специального назначения (ASIC), или программируемой логической интегральной схемы (FPGA), или же комбинации аппаратного и программного обеспечения, например с помощью микропроцессорной системы и набора инструкций, реализующих функционал модуля, которые (при выполнении) трансформируют микропроцессорную систему в устройство специального назначения. Модуль определения связи между фрагментами 112 также может быть реализован в виде комбинации аппаратного и программного обеспечения, причем некоторые конкретные функции реализуются за счет аппаратного обеспечения, а другие функции реализуются комбинацией аппаратного и программного обеспечения.

45 [00124] В других вариантах осуществления настоящей технологии модуль определения связи между фрагментами 112 может также быть реализован как программное обеспечение для выполнения различными типами процессоров. Модуль определения связи между фрагментами 112 может, например, включать один или несколько

физических или логических блоков компьютерных инструкций, которые могут быть организованы, например, как объект, процедура или функция. Однако данные исполнимые модули не должны быть физически расположены вместе и они могут включать в себя различные инструкции, сохраненные в различных местах, которые, при логическом соединении друг с другом, включают в себя данный модуль и выполняют заявленную задачу для данного модуля.

[00125] Модуль определения связи между фрагментами 112 выполнен с возможностью осуществлять определение связи между каждой парой фрагментов из множества фрагментов, по меньшей мере на основе синтаксического (парсинг) и семантического анализов текстового представления цифрового контента.

[00126] В некоторых вариантах осуществления настоящей технологии модуль определения связи между фрагментами 112 выполнен с возможностью осуществлять определение связи между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно обученного третьего алгоритма машинного обучения. В некоторых вариантах осуществления настоящей технологии третий алгоритм машинного обучения предварительно обучен осуществлять определение связи между фрагментами, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные синтаксического и семантического анализов текстового представления каждого фрагмента цифрового контента, а обучение третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

[00127] В некоторых вариантах осуществления настоящей технологии, в которых цифровой документ 302 представляет собой аудио- или видеофайл, модуль определения связи между фрагментами 112 выполнен с возможностью выполнять дополнительный анализ звукового представления цифрового контента на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы. Причем третий алгоритм машинного обучения дополнительно предварительно обучают определять связь между каждой парой фрагментов из множества фрагментов на основе анализа звукового представления цифрового контента.

[00128] Например, для случая, когда цифровой документ 302 представляет собой аудиокнигу, определенную роль при определении связи между фрагментами из множества фрагментов цифрового контента может играть звуковое представление цифрового контента, в частности голос диктора, музыкальное сопровождение, интонация диктора, повышение тона, наличие пауз и их длительность и др.

[00129] Например, для случая, когда цифровой документ 302 представляет собой видеолекцию, определенную роль при определении связи между каждыми двумя фрагментами из множества фрагментов может играть сменяемость голосов (диалог), интонация лектора. Для случая, когда цифровой документ 302 представляет собой фильм, при определении связи между фрагментами определенную роль может играть музыкальное сопровождение, голоса актеров, их тембр, тон, интонация. Например, повторяемость музыкального сопровождения и/или повторяемость голосов может свидетельствовать о продолжении сюжетной линии в фильме и связанности фрагментов между собой.

[00130] В некоторых вариантах осуществления настоящей технологии, в которых цифровой документ 302 представляет собой видеофайл, модуль определения связи между фрагментами 112 выполнен с возможностью осуществлять дополнительный анализ визуального представления цифрового контента на основе по меньшей мере одного из следующего: фон, сцена, цветность, образы и др. Причем третий алгоритм

машинного обучения дополнительно предварительно обучают определять связь между каждой парой фрагментов из множества фрагментов на основе анализа визуального представления цифрового контента.

5 [00131] Возможен вариант осуществления настоящей технологии, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра третьего алгоритма машинного обучения. Так процессор 104 посредством модуля определения связи между фрагментами
10 выполнен с возможностью отправки по меньшей мере одной пары фрагментов из множества фрагментов цифрового документа 302 человеку-ассессору для оценки связи между этой парой фрагментов, например между вторым фрагментом 3022 и первым фрагментом 3021. Оценка, полученная от ассессора, может быть использована в качестве дополнительного входного параметра третьего алгоритма машинного обучения для
15 определения связи между каждой парой фрагментов из множества фрагментов цифрового документа 302 и/или другого цифрового контента.

[00132] Например, для случая, когда цифровой документ 302 представляет собой фильм определенную роль при определении связи между фрагментами из множества фрагментов может играть визуальное представление цифрового контента, в частности
20 повторяемость образов или сцен. Например, изображение одного и того же объекта в кадре может свидетельствовать о наличии связи между фрагментами.

[00133] Стоит отметить, что приведенные выше примеры с аудиокнигой, видеолекцией и фильмом носят лишь иллюстративный характер и не являются ограничивающими для настоящей технологии, а предназначены лишь для облегчения понимания. Первый,
25 второй и третий алгоритмы машинного обучения могут быть обучены определять категорию контента и учитывать различные параметры, характерные для данной категории контента. При этом одни и те же характеристики для разных категорий контента могут означать противоположное. Например, для фильма повторяемость сцены скорее всего будет означать связанность фрагментов (продолжение сюжетной
30 линии), в то время как для видеолекции будет характерно незначительное изменение сцен, и их повторяемость не обязательно будет означать связь между фрагментами. Для фильма наиболее динамичные сцены с громким звуковым сопровождением могут означать высокую полезность фрагмента, а для видеолекции наоборот динамичная сцена и наличие шумов может означать низкую полезность (студенты заходят в
35 аудиторию, лектор настраивает громкость микрофона, вытирает доску и т.д.).

[00134] На Фиг. 2 представлена блок-схема способа 200, выполняемого в рамках системы 100, изображенной на Фиг. 1, и выполняемого в соответствии с вариантом осуществления настоящей технологии.

[00135] Этап 202 - получение указания на цифровой контент

40 [00136] Способ 200 начинается на этапе 202, на котором компьютерное устройство 102 получает указание на цифровой документ 302 от сервера источника цифрового контента 130.

[00137] Возможен вариант осуществления настоящей технологии, в котором компьютерное устройство 102 получает указание на цифровой документ 302 от
45 клиентского устройства 140.

[00138] Если цифровой документ 302 представляет собой аудиофайл, видеофайл или изображение, то модуль получения цифрового контента 106 дополнительно выполняет получение текстового представления цифрового документа 302 от сервера источника

цифрового контента 130 или от клиентского устройства 140. Если текстовое представление цифрового документа 302 недоступно, то модуль получения цифрового контента 106 выполняет создание текстового представления цифрового документа 302.

[00139] Затем способ продолжается на этапе 204.

5 [00140] Этап 204 – синтаксический анализ текстового представления цифрового контента

[00141] На этапе 204 компьютерное устройство 102 выполняет синтаксический анализ (парсинг) текстового представления цифрового документа 302 посредством модуля разделения цифрового контента 108.

10 [00142] Возможен вариант осуществления способа, в котором модуль разделения цифрового контента 108 выполняет анализ звукового и/или визуального представления цифрового документа 302. При этом различные звуковые и визуальные эффекты размечаются в соответствующем текстовом представлении цифрового документа 302. Например, для случая, когда цифровой документ 302 представляет собой видеолекцию, 15 фрагмент, на котором лектор замедляет темп речи, соотносится с предложением, которое он произносит и размечается в текстовом представлении. Аналогичным образом изменение сцены, например переход с общего плана к изображению доски, также соотносится с текстовым и/или звуковым представлением, после чего размечается в текстовом представлении цифрового документа. В качестве неограничивающего примера 20 звуковых эффектов можно назвать следующие: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы. В качестве неограничивающего примера визуальных эффектов можно назвать следующие: фон, сцена, цветность.

[00143] Затем способ продолжается на этапе 206.

25 [00144] Этап 206 – разделение цифрового контента на упорядоченное множество фрагментов, включающее в себя первый фрагмент и второй фрагмент

[00145] Этап 206 будет описан со ссылкой на Фиг. 2 и Фиг 3. На этапе 206 компьютерное устройство 102 посредством модуля разделения цифрового контента 108 и на основе синтаксического анализа текстового представления цифрового документа 302 выполняет разделение цифрового документа 302 на упорядоченное 30 множество фрагментов, включающее в себя первый фрагмент 3021 и второй фрагмент 3022.

[00146] Возможен вариант осуществления настоящей технологии, в котором на этапе 206 разделение цифрового документа 302 на упорядоченное множество фрагментов дополнительно выполняют на основе анализа звукового и/или визуального 35 представления цифрового документа 302.

[00147] Возможен вариант осуществления настоящей технологии, в котором разделение цифрового контента на упорядоченное множество фрагментов осуществляют по меньшей мере частично на основе предварительно обученного первого алгоритма машинного обучения. Дополнительно разделение цифрового документа 302 могут 40 осуществлять с участием ассессоров в целях проверки и дообучения первого алгоритма машинного обучения.

[00148] Возможен вариант осуществления настоящей технологии, в котором размер первого фрагмента 3021 не равен размеру второго фрагмента 3022.

45 [00149] Возможен вариант осуществления настоящей технологии, в котором упорядоченное множество фрагментов включает в себя по меньшей мере три фрагмента: первый 3021, второй 3022 и третий 3023.

[00150] Как показано в иллюстративном примере на Фиг.3, цифровой документ 302 разделен на упорядоченное множество фрагментов, включающее в себя восемь

фрагментов (3021, 3022, 3023, 3024, 3025, 3026, 3027, 3028). Однако цифровой контент в контексте настоящей технологии может быть разделен на большее или меньшее количество фрагментов.

[00151] Затем способ продолжается на этапе 208.

5 [00152] Этап 208 – семантический анализ каждого фрагмента текстового представления цифрового контента

[00153] На этапе 208 компьютерное устройство 102 посредством модуля определения параметра полезности 110 выполняет семантический (смысловой анализ) каждого фрагмента текстового представления цифрового документа 302. Для осуществления семантического анализа могут быть применены различные алгоритмы, например алгоритм нахождения семантических зависимостей с помощью базовых семантических шаблонов, алгоритмы, основанные на правилах, которые описывают все возможные варианты использования той или иной лексической единицы в тексте; алгоритмы, основанные на статистике (с применением машинного обучения); смешанные алгоритмы, в которых используются как статистические методологии, так и словари и др.

[00154] Затем способ продолжается на этапе 210.

[00155] Этап 210 – определение параметра полезности для каждого фрагмента из множества фрагментов

[00156] Этап 210 будет описан со ссылкой на Фиг. 4. На этапе 210 компьютерное устройство 102 посредством модуля определения параметра полезности 110 и на основе результатов семантического анализа выполняет определение параметра полезности каждого фрагмента из упорядоченного множества фрагментов цифрового документа 302.

[00157] Возможен вариант осуществления настоящей технологии, в котором параметр полезности каждого фрагмента из множества фрагментов определяют посредством по меньшей мере частично на основе предварительно обученного второго алгоритма машинного обучения. Второй алгоритм машинного обучения предварительно обучен определять параметр полезности каждого фрагмента из множества фрагментов, причем в качестве входных параметров второго алгоритма машинного обучения задают данные семантического анализа текстового представления цифрового документа 302, а обучение второго алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

[00158] Дополнительно определение параметра полезности по меньшей мере одного из фрагментов цифрового документа 302 могут осуществлять с участием ассессоров в целях проверки и дообучения второго алгоритма машинного обучения. Так, например, множеству ассессоров может быть показан по меньшей мере один фрагмент из множества фрагментов (3021-3028) цифрового документа 302. После чего среднюю оценку полезности для по меньшей мере одного фрагмента, указанную ассессорами, сравнивают с оценкой, полученной посредством второго алгоритма машинного обучения.

[00159] Так параметр полезности может быть определен для каждого фрагмента из множества фрагментов 3021-3028 цифрового документа 302. Как показано на фиг. 4, параметр полезности каждого из фрагментов 3021- 3028 определен в диапазоне от 0 до 1. Причем наибольшая условная полезность для пользователя соответствует 1, а наименьшая – 0. В иллюстративном примере на фиг. 4 наибольшей полезностью обладает второй фрагмент 3022 и седьмой фрагмент 3027. При этом, первый фрагмент 3021 обладает наименьшей полезностью среди всех остальных фрагментов цифрового документа 302.

[00160] Возможен вариант осуществления технологии, в котором в ответ на получение

запроса на создание цифрового контента отправляют ассессору по меньшей мере одно указание на фрагмент цифрового контента для оценки его параметра полезности; на этапе определения параметра полезности для каждого фрагмента из множества фрагментов в качестве дополнительного входного параметра используют упомянутую
5 полученную оценку параметра полезности по меньшей мере одного из фрагментов. При получении запроса на создание краткого изложения цифрового контента от клиентского устройства 140 возможен вариант осуществления, в котором в качестве ассессора используют пользователя 150, отправившего запрос с клиентского устройства 140. Таким образом, оценка параметра полезности по меньшей мере одного из
10 фрагментов может быть получена от пользователя 150, то есть полезность остальных фрагментов будет определена с учетом оценки полезности для конкретного пользователя 150.

[00161] Возможен вариант осуществления настоящей технологии, в котором запрос на создание краткого изложения цифрового контента включает в себя пороговое
15 значение параметра полезности фрагментов цифрового контента для включения в краткое изложение цифрового контента. Пороговое значение может быть задано заранее перед осуществлением способа. На Фиг. 4 показано пороговое значение 402 и дополнительное пороговое значение 404. В зависимости от величины порогового значения в краткое изложение цифрового контента может быть включено разное
20 количество фрагментов цифрового документа 302.

[00162] Затем способ продолжается на этапе 212.

[00163] Этап 212 – определение связи между каждой парой фрагментов из множества фрагментов

[00164] Этап 212 будет описан со ссылкой на фиг. 5. На этапе 212 процессор 104
25 компьютерного устройства 102 посредством модуля определения связи 112 и на основе семантического и синтаксического анализа текстового представления цифрового документа 302 выполняет определение связи между каждой парой фрагментов из множества фрагментов 3021-3028 цифрового документа 302. Под определением связи между каждой парой понимается определение связи между первым фрагментом 3021
30 и вторым фрагментом 3022; первым фрагментом 3021 и третьим фрагментом 3023, ..., вторым фрагментом 3022 и первым фрагментом 3021, вторым фрагментом 3022 и третьим фрагментом 3023 и т.д. То есть сравнение каждого фрагмента с каждым.

[00165] Возможен вариант осуществления настоящей технологии, в котором связь между каждой парой фрагментов из множества фрагментов дополнительно определяют
35 на основе анализа звукового и/или визуального представления цифрового документа 302.

[00166] Возможен вариант осуществления настоящей технологии, в котором определяют связь между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно обученного третьего алгоритма
40 машинного обучения.

[00167] Возможен вариант осуществления настоящей технологии, в котором третий алгоритм машинного обучения предварительно обучен определять связь между каждой парой фрагментов из множества фрагментов, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные семантического и
45 синтаксического (парсинга) анализов текстового представления цифрового контента, а обучение третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки. В качестве дополнительных входных параметров третьего алгоритма машинного обучения могут быть использованы данные

анализа звукового и/или визуального представления цифрового документа 302.

[00168] На фиг. 5 согласно неограничивающему иллюстративному примеру осуществления настоящей технологии схематично показаны связи между фрагментами 3021-3028 цифрового документа 302. Под термином «взаимосвязь фрагментов», «связанность фрагментов», «связь фрагментов» в контексте настоящей технологии понимается понятность для пользователя одного фрагмента при исключении другого фрагмента. Так в контексте фиг. 5 однонаправленной стрелкой обозначено то, что фрагмент, от которого исходит стрелка, связан с (не понятен без фрагмента, к которому направлена стрелка) фрагментом, к которому направлена стрелка. А двунаправленная стрелка обозначает то, что фрагменты не понятны друг без друга, то есть между ними присутствует взаимосвязь. Отсутствие стрелок между фрагментами обозначает отсутствие связи, то есть фрагменты являются понятными пользователю по отдельности. Посредством предварительно обученного третьего алгоритма машинного обучения могут быть определены связи между каждыми двумя фрагментами из множества фрагментов. На наличие связи между фрагментами может влиять, например, первое и многократное упоминание термина в одном фрагменте, и менее частое упоминание данного термина в другом фрагменте может свидетельствовать, например, о том, что в одном фрагменте дано подробное определение термина и, следовательно, другой фрагмент может быть связан с ним (не понятен без него).

[00169] Как показано на Фиг. 5, между первым фрагментом 3021 и вторым фрагментом 3022 присутствует взаимосвязь. Возможен вариант осуществления настоящей технологии, в котором в ответ на получение указания на взаимосвязь между первым 3021 и вторым 3022 фрагментом выполняют объединение первого 3021 и второго 3022 фрагмента в единый фрагмент 3030, как это показано на фиг. 3.

[00170] Возможен вариант осуществления настоящей технологии, в котором определение связи между каждой парой фрагментов выполняют параллельно с этапом разделения цифрового контента на упорядоченное множество фрагментов.

[00171] Возможен вариант осуществления настоящей технологии, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра третьего алгоритма машинного обучения.

[00172] Возможен вариант осуществления настоящей технологии, в котором указание на связь между каждыми двумя фрагментами из множества фрагментов представляет собой: параметр, указывающий на наличие связи или взаимосвязи между первым и вторым фрагментом, либо параметр, указывающий степень связи или взаимосвязи между первым и вторым фрагментом. Так, например, второй фрагмент 3022 не понятен без первого фрагмента 3021 для 60% пользователей. Следовательно, в таком примере степень взаимосвязи будет составлять 60%. Для оставшихся 40% второй фрагмент 3022 может быть понятен без первого фрагмента 3021. В качестве иллюстративного примера, для случая видеолекции второй фрагмент 3022 может быть не понятен большинству без вводной информации, которая дается в первом фрагменте 3021. Для многих других, кому известна информация из первого фрагмента 3021, второй фрагмент 3022 может быть понятен без первого фрагмента 3021.

[00173] Возможен вариант осуществления настоящей технологии, в котором указание на взаимосвязь представляет собой параметр, указывающий степень связи или взаимосвязи между фрагментами; причем перед проверкой связи между каждой парой

фрагментов из подмножества фрагментов задают пороговое значение параметра, указывающего степень связи или взаимосвязи между фрагментами. Например, для случая, когда краткое изложение цифрового контента рассчитано на продвинутого пользователя, то пороговое значение параметра связи/взаимосвязи может быть

5 увеличено по сравнению пороговым значением параметра связи, взаимосвязи в кратком изложении цифрового контента, которая рассчитана на более широкую аудиторию (для менее продвинутых пользователей). Возвращаясь к примеру с цифровым документом 302, который представляет собой видеолекцию, краткое изложение цифрового документа 302 может быть предназначено для студентов и школьников.

10 Альтернативно, краткое изложение цифрового документа 302 может быть предназначено для более продвинутых пользователей, например ученых, владеющих вводной информацией из первого фрагмента 3021, таким пользователям второй фрагмент 3022 будет понятен без первого фрагмента 3021.

[00174] Возможен вариант осуществления, в котором запрос на создание краткого

15 изложения цифрового контента включает в себя указание на степень владения тематикой цифрового контента. При получении запроса на создание краткого изложения цифрового контента от клиентского устройства 140, возможен вариант осуществления, в котором пользователь 150 может задавать степень владения тематикой цифрового документа 302, для которого необходимо создать краткое изложение. Под степенью владения

20 тематикой может пониматься, например, два крайних значения, условно: «новичок» и «продвинутый». Как будет понятно специалистам, условная степень владения тематикой может быть оценена иным образом, например в баллах от 1 до 5, или иначе.

[00175] Возможен вариант осуществления, в котором на этапе определения связи между фрагментами в качестве входного параметра используют полученное значение

25 степени владения тематикой цифрового контента. С учетом полученной от пользователя 150 условной оценки владения тематикой цифрового документа 302 значение наличия связи между фрагментами или степени связи между фрагментами могут быть скорректированы. Например, цифровой документ 302 представляет собой фильм, при этом пользователь 150 читал книгу, по которой снят фильм, и в качестве степени

30 владения тематикой выбрал «продвинутый». При этом часть фрагментов цифрового контента, которые не обладают высокой полезностью, но с которыми связаны наиболее полезные фрагменты, смогут быть исключены из краткого изложения цифрового документа 302, поскольку пользователь 150 хорошо знаком с тематикой цифрового документа 302 и он сможет понять некоторые фрагменты, обладающие параметром

35 полезности выше порогового значения, которые связаны с другими фрагментами, обладающими параметром полезности ниже порогового значения. То есть «продвинутый» пользователь 150 сможет достроить связи и понять, например, второй фрагмент 3022 без первого фрагмента 3021, в то время как пользователь «новичок» не сможет самостоятельно достроить связи и следовательно он не сможет понять второй

40 фрагмент 3022 без первого фрагмента 3021. Таким образом, может быть достигнут баланс между высокой полезностью, высокой связанностью (понятностью) краткого изложения цифрового контента и размером краткого изложения цифрового контента с учетом конкретного пользователя 150.

[00176] Возможен вариант осуществления настоящей технологии, в котором на

45 основе полученного значения степени владения тематикой цифрового контента определяют параметр полезности по меньшей мере одного фрагмента цифрового контента. Так, например, один и тот же фрагмент цифрового контента может обладать высокой полезностью для пользователя «новичка» и низкой полезностью для

«продвинутого» пользователя.

[00177] Затем способ продолжается на этапе 214.

[00178] Этап 214 – в ответ на то, что параметр полезности второго фрагмента превышает предварительно определенное пороговое значение параметра полезности, включение второго фрагмента в подмножество фрагментов для включения в краткое изложение цифрового контента;

[00179] На этапе 214 процессор 104 компьютерного устройства 102 создает подмножество фрагментов, в которое включают фрагменты, параметры полезности которых превышают предварительно определенное пороговое значение.

[00180] На фиг. 4 показано пороговое значения полезности 402, которое составляет 0,5. Следовательно, все фрагменты, параметр полезности которых превышает пороговое значение 402, включаются в подмножество фрагментов для включения в краткое изложение цифрового документа 302.

[00181] Возможен вариант осуществления настоящей технологии, в котором перед определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение 404 параметра полезности; создают по меньшей мере одно дополнительное подмножество фрагментов, на основе которого создают по меньшей мере одно дополнительное краткое изложение цифрового контента.

[00182] Возможен вариант осуществления настоящей технологии, в котором по меньшей мере одно дополнительное краткое изложение цифрового контента отличается по размеру от краткого изложения цифрового контента.

[00183] Затем способ продолжается на этапе 216.

[00184] Этап 216 – в ответ на получение указания на связь второго фрагмента с первым фрагментом включение первого фрагмента в подмножество фрагментов

[00185] На этапе 216 в ответ на получение указания на связь второго фрагмента 3022 с первым 3021 фрагментом включают первый фрагмент 3021 в подмножество фрагментов.

[00186] В представленном иллюстративном примере параметр полезности первого фрагмента 3021 не превышает предварительно определенное пороговое значение (402, 404). Несмотря на это он включается в подмножество фрагментов для включения в краткое изложение цифрового документа 302.

[00187] Затем способ продолжается на этапе 218.

[00188] Этап 218 – создание краткого изложения цифрового контента на основе подмножества фрагментов

[00189] Со ссылкой на фиг. 6. На этапе 218 создают краткое изложение 602 цифрового документа 302 на основе подмножества фрагментов, включающего в себя первый фрагмент 3021, второй фрагмент 3022, пятый фрагмент 3025, седьмой фрагмент 3027 и восьмой фрагмент 3028.

[00190] Возможен вариант осуществления настоящей технологии, в котором перед определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение 404 параметра полезности; на этапе 218 создания подмножества фрагментов создают по меньшей мере одно дополнительное подмножество фрагментов, на основе которого создают по меньшей мере одно дополнительное краткое изложение 604 цифрового контента. Дополнительное краткое изложение 604 цифрового контента согласно иллюстративному примеру на фиг. 6 включает в себя первый фрагмент 3021, второй фрагмент 3022 и седьмой фрагмент 3027.

[00191] Дополнительное краткое изложение 604 цифрового контента отличается по размеру от краткого изложения 602 цифрового контента, поскольку дополнительное пороговое значение параметра полезности 404 превышает пороговое значение 402 параметра полезности и, следовательно, меньшее число фрагментов было включено в
5 дополнительное подмножество фрагментов для создания дополнительного краткого изложения 604 цифрового документа 302.

[00192] Стоит отметить, что первый фрагмент 3021 был включен в краткое изложение 602 цифрового документа 302 и дополнительное краткое изложение 604 цифрового контента несмотря на то, что параметр полезности первого фрагмента 3021 не
10 превышает пороговое значение 402 параметра полезности и дополнительное пороговое значение параметра полезности 404. Это обусловлено тем, что между вторым фрагментом 3022 и первым фрагментом 3022 имеется взаимосвязь.

[00193] После создания краткого изложения цифрового контента 602 и 604 они могут быть сохранены, в частности, на сервере источника цифрового контента 130, или на
15 компьютерном устройстве 102, или отправлены на клиентское устройство 140.

[00194] После этапа 218 способ 200 может завершаться.

[00195] Передача-получение различных сигналов, в том числе задач и запросов, хорошо известны в данной области техники и поэтому для упрощения были опущены в конкретных частях данного описания. Сигналы могут быть переданы-получены с
20 помощью оптических средств (например, оптоволоконного соединения), электронных средств (например, проводного или беспроводного соединения) и механических средств (например, на основе давления, температуры или другого подходящего параметра).

[00196] Таким образом с некоторой точки зрения, способ и система, реализованные в соответствии с некоторыми неограничивающими вариантами осуществления
25 настоящей технологии, могут быть представлены следующим образом, представленными в пронумерованных пунктах.

[00197] ПУНКТ 1. Способ (200) создания краткого изложения цифрового контента (302), осуществляемый процессором 104 компьютерного устройства 102, способ (200) включает в себя:

30 получение (202) указания на цифровой контент (302);
синтаксический анализ (204) текстового представления цифрового контента (302);
разделение (206) цифрового контента на упорядоченное множество фрагментов, включающее в себя первый фрагмент 3021 и второй фрагмент 3022;

35 семантический анализ (208) каждого фрагмента текстового представления цифрового контента (302);

определение (210) параметра полезности для каждого фрагмента из множества фрагментов;

40 определение (212) связи между каждой парой фрагментов из множества фрагментов;
(214) в ответ на то, что параметр полезности второго фрагмента превышает
предварительно определенное пороговое значение параметра полезности, включение второго фрагмента (3022) в подмножество фрагментов для включения в краткое изложение цифрового контента;

(216) в ответ на получение указания на связь второго фрагмента (3022) с первым фрагментом (3021) включение первого фрагмента (3021) в подмножество фрагментов;
45 создание (218) краткого изложения цифрового контента (302) на основе подмножества фрагментов.

[00198] ПУНКТ 2. Способ (200) по п. 1, в котором разделение цифрового контента на упорядоченное множество фрагментов осуществляют по меньшей мере частично

на основе предварительно обученного первого алгоритма машинного обучения.

5 [00199] ПУНКТ 3. Способ (200) по п. 2, в котором первый алгоритм машинного обучения предварительно обучен разделять цифровой контент на множество фрагментов, причем в качестве входных параметров первого алгоритма машинного обучения задают данные синтаксического анализа текстового представления цифрового контента, а обучение первого алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

10 [00200] ПУНКТ 4. Способ (200) по любому из пп. 1- 3, в котором параметр полезности каждого фрагмента из множества фрагментов определяют по меньшей мере частично на основе предварительно обученного второго алгоритма машинного обучения.

15 [00201] ПУНКТ 5. Способ (200) по п.4, в котором второй алгоритм машинного обучения предварительно обучен определять параметр полезности каждого фрагмента из множества фрагментов, причем в качестве входных параметров второго алгоритма машинного обучения задают данные семантического анализа текстового представления цифрового контента, а обучение второго алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

20 [00202] ПУНКТ 6. Способ (200) по п. 5, в котором на этапе (210) определения параметра полезности полезность по меньшей мере одного фрагмента из множества фрагментов оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра второго алгоритма машинного обучения.

25 [00203] ПУНКТ 7. Способ (200) по п.1, в котором указание на связь между каждыми двумя фрагментами из множества фрагментов представляет собой: параметр, указывающий на наличие связи между каждыми двумя фрагментами, либо параметр, указывающий на степень связанности между каждыми двумя фрагментами.

[00204] ПУНКТ 8. Способ (200) по п. 1, в котором определяют связь между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно обученного третьего алгоритма машинного обучения.

30 [00205] ПУНКТ 9. Способ (200) по п. 8, в котором третий алгоритм машинного обучения предварительно обучен определять связь между каждой парой фрагментов из множества фрагментов, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные семантического и синтаксического анализов текстового представления цифрового контента, а обучение третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

40 [00206] ПУНКТ 10. Способ (200) по п. 9, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра третьего алгоритма машинного обучения.

[00207] ПУНКТ 11. Способ (200) по любому из пп. 1-10, в котором параметр полезности первого фрагмента не превышает предварительно определенное пороговое значение.

45 [00208] ПУНКТ 12. Способ (200) по любому из пп. 1-11, в котором перед (210) определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение параметра полезности; создают по меньшей мере одно дополнительное краткое изложение цифрового контента на основе по меньшей мере одного дополнительного подмножества

фрагментов.

[00209] ПУНКТ 13. Способ (200) по п. 12, в котором по меньшей мере одно дополнительное краткое изложение цифрового контента отличается по размеру от краткого изложения цифрового контента.

5 [00210] ПУНКТ 14. Способ (200) по любому из пп. 1-13, в котором перед этапом (206) разделения цифрового контента получают запрос на создание краткого изложения цифрового контента.

10 [00211] ПУНКТ 15. Компьютерное устройство (102) для создания краткого изложения цифрового контента, включающее в себя процессор (104), коммуникационный интерфейс, процессор (104) функционально связан с модулем получения цифрового контента (106), модулем разделения цифрового контента на упорядоченное множество фрагментов (108), модулем определения параметра полезности (110), модулем определения связи между фрагментами (112), причем процессор (104) выполнен с возможностью осуществлять способ (200) по пп. 1-14.

15 [00212] Модификации и улучшения вышеописанных вариантов осуществления настоящей технологии будут ясны специалистам в данной области техники. Предшествующее описание представлено только в качестве примера и не несет никаких ограничений. Таким образом, объем настоящей технологии ограничен только объемом прилагаемой формулы изобретения.

20

(57) Формула изобретения

1. Способ создания краткого изложения цифрового контента, осуществляемый процессором компьютерного устройства, способ включает в себя:

получение указания на цифровой контент;

25 синтаксический анализ текстового представления цифрового контента;

разделение цифрового контента на упорядоченное множество фрагментов, включающее в себя первый фрагмент и второй фрагмент;

семантический анализ каждого фрагмента текстового представления цифрового контента;

30 определение параметра полезности для каждого фрагмента из множества фрагментов;

определение связи между каждой парой фрагментов из множества фрагментов;

в ответ на то, что параметр полезности второго фрагмента превышает предварительно определенное пороговое значение параметра полезности, включение второго фрагмента в подмножество фрагментов для включения в краткое изложение цифрового контента;

35 в ответ на получение указания на связь второго фрагмента с первым фрагментом включение первого фрагмента в подмножество фрагментов;

создание краткого изложения цифрового контента на основе подмножества фрагментов.

40 2. Способ по п. 1, в котором разделение цифрового контента на упорядоченное множество фрагментов осуществляют по меньшей мере частично на основе предварительно обученного первого алгоритма машинного обучения.

3. Способ по п. 2, в котором первый алгоритм машинного обучения предварительно обучен разделять цифровой контент на множество фрагментов, причем в качестве входных параметров первого алгоритма машинного обучения задают данные синтаксического анализа текстового представления цифрового контента, а обучение первого алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

4. Способ по п. 1, в котором параметр полезности каждого фрагмента из множества фрагментов определяют по меньшей мере частично на основе предварительно обученного второго алгоритма машинного обучения.

5 5. Способ по п. 4, в котором второй алгоритм машинного обучения предварительно обучен определять параметр полезности каждого фрагмента из множества фрагментов, причем в качестве входных параметров второго алгоритма машинного обучения задают данные семантического анализа текстового представления цифрового контента, а обучение второго алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

10 6. Способ по п. 5, в котором на этапе определения параметра полезности полезность по меньшей мере одного фрагмента из множества фрагментов оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра второго алгоритма машинного обучения.

15 7. Способ по п. 1, в котором размер первого фрагмента не равен размеру второго фрагмента.

8. Способ по п. 1, в котором в ответ на получение указания на взаимосвязь второго фрагмента с первым фрагментом выполняют объединение первого и второго фрагмента в единый фрагмент.

20 9. Способ по п. 1, в котором упорядоченное множество фрагментов включает в себя по меньшей мере три фрагмента: первый, второй и третий.

10. Способ по п. 1, в котором указание на связь между каждыми двумя фрагментами из множества фрагментов представляет собой: параметр, указывающий на наличие связи между каждыми двумя фрагментами, либо параметр, указывающий на степень связанности между каждыми двумя фрагментами.

25 11. Способ по п. 1, в котором определяют связь между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно обученного третьего алгоритма машинного обучения.

30 12. Способ по п. 11, в котором третий алгоритм машинного обучения предварительно обучен определять связь между каждой парой фрагментов из множества фрагментов, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные семантического и синтаксического анализов текстового представления цифрового контента, а обучение третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

35 13. Способ по п. 12, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра третьего алгоритма машинного обучения.

40 14. Способ по п. 1, в котором определение связи между каждой парой фрагментов выполняют параллельно с этапом разделения цифрового контента на упорядоченное множество фрагментов.

45 15. Способ по п. 1, в котором перед созданием краткого изложения цифрового контента на основе подмножества фрагментов дополнительно осуществляют проверку связи между каждой парой фрагментов из подмножества фрагментов на основе предварительно обученного третьего алгоритма машинного обучения и/или на основе проверки по меньшей мере одним ассессором.

16. Способ по п. 15, в котором указание на связь представляет собой параметр, указывающий степень связанности между фрагментами; причем перед проверкой связи

между каждой парой фрагментов из подмножества фрагментов задают пороговое значение параметра, указывающего степень связанности между фрагментами.

17. Способ по п. 1, в котором параметр полезности первого фрагмента не превышает предварительно определенное пороговое значение.

5 18. Способ по п. 1, в котором перед определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение параметра полезности; создают по меньшей мере одно дополнительное краткое изложение цифрового контента на основе по меньшей мере одного дополнительного подмножества фрагментов.

10 19. Способ по п. 18, в котором по меньшей мере одно дополнительное краткое изложение цифрового контента отличается по размеру от краткого изложения цифрового контента.

20. Способ по п. 1, в котором перед этапом разделения цифрового контента получают запрос на создание краткого изложения цифрового контента.

15 21. Способ по п. 20, в котором в ответ на получение запроса на создание цифрового контента отправляют ассессору по меньшей мере одно указание на фрагмент цифрового контента для оценки его параметра полезности; на этапе определения параметра полезности для каждого фрагмента из множества фрагментов в качестве дополнительного входного параметра используют упомянутую полученную оценку
20 параметра полезности по меньшей мере одного из фрагментов.

22. Способ по п. 20, в котором запрос на создание краткого изложения цифрового контента включает в себя пороговое значение параметра полезности фрагментов цифрового контента для включения в краткое изложение цифрового контента.

23. Способ по п. 20, в котором запрос на создание краткого изложения цифрового
25 контента включает в себя указание на степень владения тематикой цифрового контента.

24. Способ по п. 23, в котором на этапе определения связи между фрагментами в качестве входного параметра используют полученное значение степени владения тематикой цифрового контента.

25. Способ по п. 1, в котором цифровой контент представляет собой текстовый файл.

30 26. Способ по п. 1, в котором цифровой контент представляет собой аудиофайл.

27. Способ по п. 26, в котором после получения цифрового контента осуществляют создание текстового представления аудиофайла.

28. Способ по п. 27, в котором при разделении аудиофайла на фрагменты и при
35 создании указаний на взаимосвязь между фрагментами дополнительно выполняют анализ на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы.

29. Способ по п. 1, в котором цифровой контент представляет собой видеофайл.

30. Способ по п. 29, в котором после получения цифрового контента осуществляют создание текстового представления видеофайла.

40 31. Способ по п. 30, в котором при разделении видеофайла на фрагменты и при определении взаимосвязи между фрагментами дополнительно выполняют анализ звукового представления на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы, а также анализ визуального представления на основе по меньшей мере одного из следующего:
45 фон, сцена, цветность.

32. Компьютерное устройство для создания краткого изложения цифрового контента, включающее в себя процессор, коммуникационный интерфейс, процессор функционально связан с модулем получения цифрового контента, модулем разделения цифрового

контента на упорядоченное множество фрагментов, модулем определения параметра полезности, модулем определения связи между фрагментами, причем процессор выполнен с возможностью осуществлять:

получение указания на цифровой контент;

5 синтаксический анализ текстового представления цифрового контента;
разделение цифрового контента на упорядоченное множество фрагментов, включающее в себя первый фрагмент и второй фрагмент;

семантический анализ каждого фрагмента текстового представления цифрового контента;

10 определение параметра полезности для каждого фрагмента из множества фрагментов;
определение связи между каждой парой фрагментов из множества фрагментов;
в ответ на то, что параметр полезности второго фрагмента превышает
предварительно определенное пороговое значение параметра полезности, включение
второго фрагмента в подмножество фрагментов для включения в краткое изложение
15 цифрового контента;

в ответ на получение указания на связь второго фрагмента с первым фрагментом
включение первого фрагмента в подмножество фрагментов;

создание краткого изложения цифрового контента на основе подмножества
фрагментов.

20 33. Компьютерное устройство по п. 32, в котором процессор выполнен с
возможностью осуществлять разделение цифрового контента на упорядоченное
множество фрагментов по меньшей мере частично на основе предварительно обученного
первого алгоритма машинного обучения.

34. Компьютерное устройство по п. 33, в котором первый алгоритм машинного
25 обучения предварительно обучен разделять цифровой контент на множество
фрагментов, причем в качестве входных параметров первого алгоритма машинного
обучения задают данные синтаксического анализа текстового представления цифрового
контента, а обучение первого алгоритма машинного обучения представляет собой
обучение с учителем на основе обучающей выборки.

30 35. Компьютерное устройство по п. 32, в котором процессор выполнен с
возможностью определять параметр полезности каждого фрагмента из множества
фрагментов по меньшей мере частично на основе предварительно обученного второго
алгоритма машинного обучения.

36. Компьютерное устройство по п. 35, в котором второй алгоритм машинного
35 обучения предварительно обучен определять параметр полезности каждого фрагмента
из множества фрагментов, причем в качестве входных параметров второго алгоритма
машинного обучения задают данные семантического анализа текстового представления
цифрового контента, а обучение второго алгоритма машинного обучения представляет
собой обучение с учителем на основе обучающей выборки.

40 37. Компьютерное устройство по п. 36, в котором на этапе определения параметра
полезности полезность по меньшей мере одного фрагмента из множества фрагментов
оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве
дополнительного входного параметра второго алгоритма машинного обучения.

38. Компьютерное устройство по п. 32, в котором размер первого фрагмента не
45 равен размеру второго фрагмента.

39 Компьютерное устройство по п. 32, в котором в ответ на получение указания на
взаимосвязь второго фрагмента с первым фрагментом процессор выполнен с
возможностью выполнять объединение первого и второго фрагмента в единый

фрагмент.

40. Компьютерное устройство по п. 32, в котором упорядоченное множество фрагментов включает в себя по меньшей мере три фрагмента: первый, второй и третий.

41. Компьютерное устройство по п. 32, в котором указание на связь между каждыми двумя фрагментами из множества фрагментов представляет собой: параметр, указывающий на наличие связи между каждыми двумя фрагментами, либо параметр, указывающий на степень связанности между каждыми двумя фрагментами.

42. Компьютерное устройство по п. 32, в котором процессор выполнен с возможностью определять связь между каждой парой фрагментов из множества фрагментов по меньшей мере частично на основе предварительно обученного третьего алгоритма машинного обучения.

43. Компьютерное устройство по п. 42, в котором третий алгоритм машинного обучения предварительно обучен определять связь между каждой парой фрагментов из множества фрагментов, причем в качестве входных параметров третьего алгоритма машинного обучения задают данные семантического и синтаксического анализов текстового представления цифрового контента, а обучение третьего алгоритма машинного обучения представляет собой обучение с учителем на основе обучающей выборки.

44. Компьютерное устройство по п. 43, в котором на этапе определения связи между каждой парой фрагментов из множества фрагментов связь между по меньшей мере одной парой фрагментов из множества фрагментов цифрового контента оценивают ассессором, причем оценку, полученную от ассессора, применяют в качестве дополнительного входного параметра третьего алгоритма машинного обучения.

45. Компьютерное устройство по п. 32, в котором процессор выполнен с возможностью определения связи между каждой парой фрагментов параллельно с этапом разделения цифрового контента на упорядоченное множество фрагментов.

46. Компьютерное устройство по п. 32, в котором перед созданием краткого изложения цифрового контента на основе подмножества фрагментов процессор выполнен с возможностью дополнительно осуществлять проверку связи между каждой парой фрагментов из подмножества фрагментов на основе предварительно обученного третьего алгоритма машинного обучения и/или на основе проверки по меньшей мере одним ассессором.

47. Компьютерное устройство по п. 46, в котором указание на связь представляет собой параметр, указывающий степень связанности между фрагментами; причем перед проверкой связи между каждой парой фрагментов из подмножества фрагментов задают пороговое значение параметра, указывающего степень связанности между фрагментами.

48. Компьютерное устройство по п. 32, в котором параметр полезности первого фрагмента не превышает предварительно определенное пороговое значение.

49. Компьютерное устройство по п. 32, в котором перед определением параметра полезности для каждого фрагмента из множества фрагментов задают по меньшей мере одно дополнительное пороговое значение параметра полезности; процессор выполнен с возможностью создавать по меньшей мере одно дополнительное краткое изложение цифрового контента на основе по меньшей мере одного дополнительного подмножества фрагментов.

50. Компьютерное устройство по п. 49, в котором по меньшей мере одно дополнительное краткое изложение цифрового контента отличается по размеру от краткого изложения цифрового контента.

51. Компьютерное устройство по п. 32, в котором процессор выполнен с

возможностью перед этапом разделения цифрового контента получать запрос на создание краткого изложения цифрового контента.

52. Компьютерное устройство по п. 51, в котором в ответ на получение запроса на создание цифрового контента процессор выполнен с возможностью отправлять ассессору по меньшей мере одно указание на фрагмент цифрового контента для оценки его параметра полезности; на этапе определения параметра полезности для каждого фрагмента из множества фрагментов в качестве дополнительного входного параметра используют упомянутую полученную оценку параметра полезности по меньшей мере одного из фрагментов.

53. Компьютерное устройство по п. 51, в котором запрос на создание краткого изложения цифрового контента включает в себя пороговое значение параметра полезности фрагментов цифрового контента для включения в краткое изложение цифрового контента.

54. Компьютерное устройство по п. 51, в котором запрос на создание краткого изложения цифрового контента включает в себя указание на степень владения тематикой цифрового контента.

55. Компьютерное устройство по п. 54, в котором на этапе определения связи между фрагментами в качестве входного параметра используют полученное значение степени владения тематикой цифрового контента.

56. Компьютерное устройство по п. 32, в котором цифровой контент представляет собой текстовый файл.

57. Компьютерное устройство по п. 32, в котором цифровой контент представляет собой аудиофайл.

58. Компьютерное устройство по п. 57, в котором после получения цифрового контента процессор выполнен с возможностью осуществлять создание текстового представления аудиофайла.

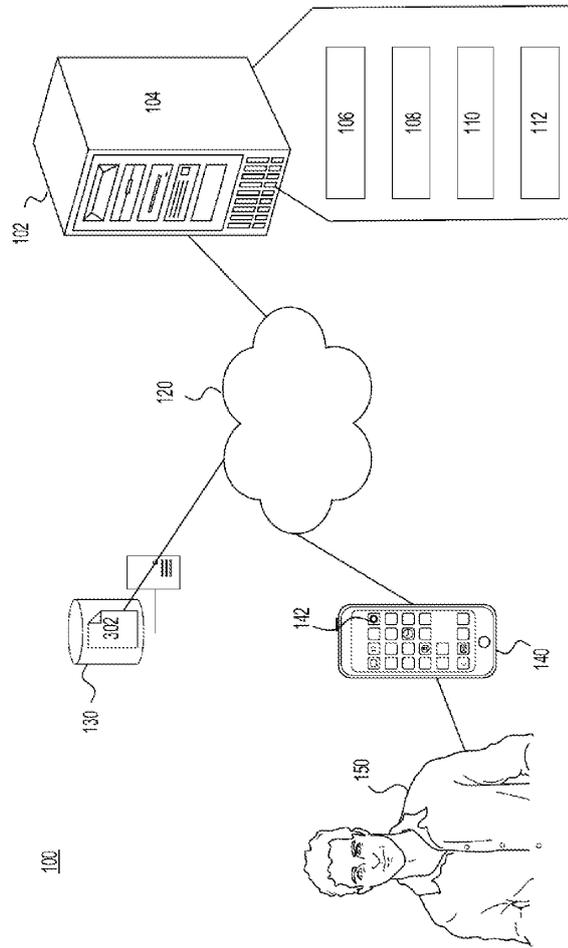
59. Компьютерное устройство по п. 58, в котором при разделении аудиофайла на фрагменты и при создании указаний на взаимосвязь между фрагментами процессор выполнен с возможностью дополнительно выполнять анализ на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы.

60. Компьютерное устройство по п. 32, в котором цифровой контент представляет собой видеофайл.

61. Компьютерное устройство по п. 60, в котором после получения цифрового контента процессор выполнен с возможностью осуществлять создание текстового представления видеофайла.

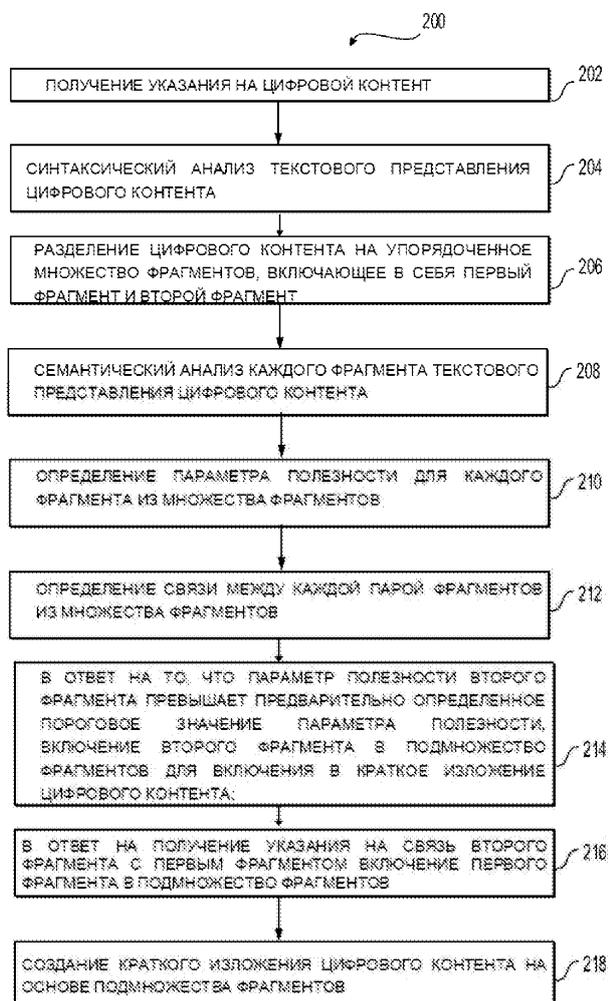
62. Компьютерное устройство по п. 61, в котором при разделении видеофайла на фрагменты и при определении взаимосвязи между фрагментами процессор выполнен с возможностью дополнительно выполнять анализ звукового представления на основе по меньшей мере одного из следующего: интонация, музыкальное сопровождение, тембр голоса, темп голоса, тон голоса, паузы, а также анализ визуального представления на основе по меньшей мере одного из следующего: фон, сцена, цветность.

1

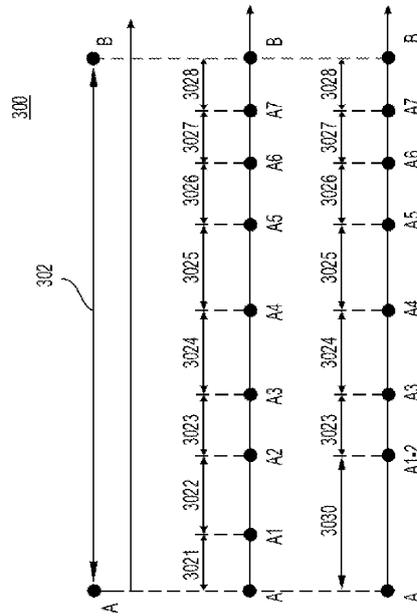


ФИГ. 1

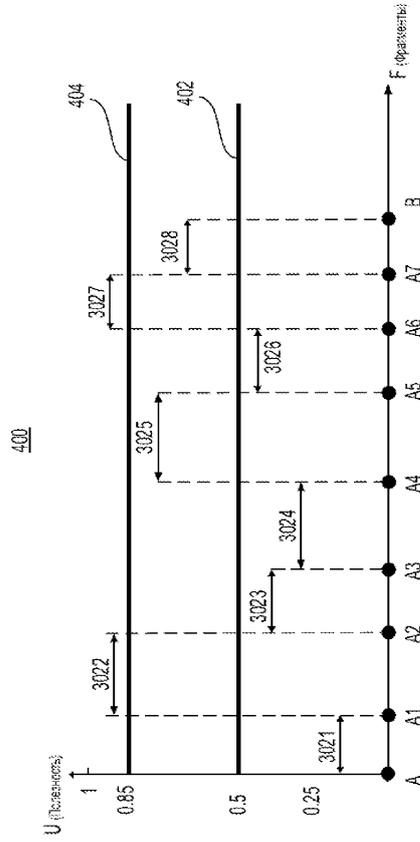
2



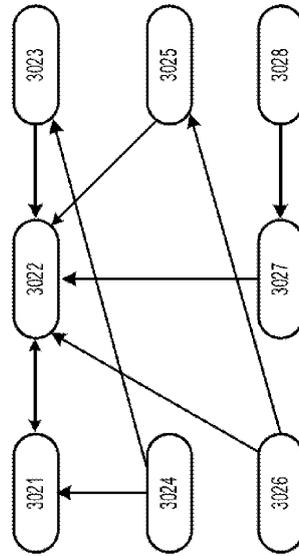
ФИГ. 2



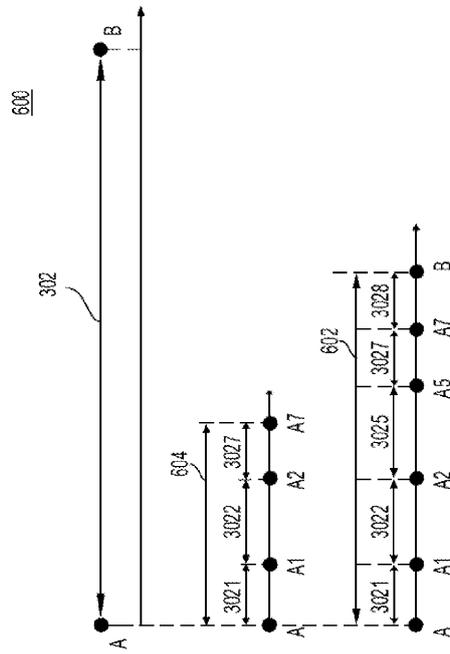
ФИГ. 3



ФИГ. 4



ФИГ. 5



ФИГ. 6