

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ЗАЯВКА НА ИЗОБРЕТЕНИЕ

(21)(22) Заявка: 2022134630, 27.12.2022

Приоритет(ы):

(22) Дата подачи заявки: 27.12.2022

(43) Дата публикации заявки: 27.06.2024 Бюл. № 18

Адрес для переписки:

119021, Москва, ул. Льва Толстого, 16, ООО
"Яндекс", патентный отдел, Малахова Надия
Ахавовна

(71) Заявитель(и):

ОБЩЕСТВО С ОГРАНИЧЕННОЙ
ОТВЕТСТВЕННОСТЬЮ "ЯНДЕКС" (RU)

(72) Автор(ы):

Кириченко Владимир Владимирович (RU),
Петренко Дарья Павловна (RU),
Панасюк Артем Андреевич (RU)

(54) СПОСОБ И СЕРВЕР ДЛЯ ГЕНЕРИРОВАНИЯ МОДИФИЦИРОВАННОГО АУДИО ДЛЯ ВИДЕО

(57) Формула изобретения

1. Способ генерирования модифицированных аудиоданных для видеофайла, причем видеофайл ассоциирован с аудиоданными, способ, исполняемый сервером, содержит этапы, на которых:

получают сервером последовательность первых аудиофрагментов,

данный фрагмент из последовательности первых аудиофрагментов представляет данное предложение на первом языке и ассоциирован с временной меткой в видеофайле;

получают сервером последовательность вторых аудиофрагментов,

данный фрагмент из последовательности вторых аудиофрагментов представляет собой другое данное предложение на втором языке и ассоциирован с временной продолжительностью, при этом другое данное предложение является переводом данного предложения;

генерируют сервером множество возможных компоновок для последовательности вторых аудиофрагментов,

данная возможная компоновка ассоциирована с возможными временными метками в видеофайле и возможными коэффициентами сжатия для соответствующих из последовательности вторых аудиофрагментов;

выбирают целевую компоновку для последовательности вторых аудиофрагментов из множества возможных компоновок, причем выбор основан на целевой компоновке, имеющей минимальную оценку потери среди оценок потери, ассоциированных с соответствующими из множества возможных компоновок, при этом данная оценка потери данной возможной компоновки генерируется на основе:

(i) комбинации различий между временными метками, ассоциированными с соответствующими из последовательности первых аудиофрагментов и соответствующими из данной возможной компоновки,

(ii) наличия перекрытия между данной парой вторых аудиофрагментов, если: вторые аудиофрагменты из последовательности вторых аудиофрагментов скомпонованы в

A
2022134630
RUR U
2 0 2 2 1 3 4 6 3 0
A

видеофайле в соответствии с возможными временными метками и возможными коэффициентами сжатия; и

генерируют по меньшей мере один модифицированный аудиофрагмент для видеофайла в качестве перевода аудиоданных с использованием целевой компоновки.

2. Способ по п. 1, в котором данный фрагмент из последовательности первых аудиофрагментов представляет собой аудио сигнал формы волны.

3. Способ по п. 1, в котором данный фрагмент из последовательности первых аудиофрагментов представляет собой мел-спектрограмму.

4. Способ по п. 1, дополнительно содержащий этап, на котором генерируют сервером последовательность вторых аудиофрагментов на основе последовательности первых аудиофрагментов.

5. Способ по п. 1, в котором временная метка, ассоциированная с данным фрагментом из последовательности первых аудиофрагментов, является временной меткой начала, а данный фрагмент из последовательности первых аудиофрагментов дополнительно ассоциирован с временной меткой окончания в видеофайле, при этом возможные временные метки данной возможной компоновки включают в себя возможные временные метки начала и возможные временные метки окончания для соответствующих вторых аудиофрагментов из данной возможной компоновки, при этом заданная оценка потери дополнительно генерируется на основе:

комбинации различий между временными метками начала, ассоциированными с соответствующими из последовательности первых аудиофрагментов, и соответствующими возможными временными метками начала из данной возможной компоновки,

комбинации различий между временными метками окончания, ассоциированными с соответствующими из последовательности первых аудиофрагментов, и соответствующими временными метками окончания из данной возможной компоновки.

6. Способ по п. 1, в котором временная метка, ассоциированная с данным фрагментом из последовательности первых аудиофрагментов, представляет собой центральную временную метку, при этом возможные временные метки данной возможной компоновки включают в себя возможные центральные временные метки для соответствующих вторых аудиофрагментов из данной возможной компоновки, при этом данная оценка потери дополнительно генерируется на основе:

комбинации различий между центральными временными метками, ассоциированными с соответствующими из последовательности первых аудиофрагментов, и соответствующими возможными центральными метками из данной возможной компоновки.

7. Способ по п. 1, в котором оценка потери дополнительно генерируется на основе комбинации возможных коэффициентов сжатия для данной возможной компоновки.

8. Способ по п. 1, в котором генерирование по меньшей мере одного модифицированного аудиофрагмента выполняется сервером в автономном режиме, и при этом способ дополнительно содержит этап, на котором сохраняют сервером видеофайл с по меньшей мере одним модифицированным аудиофрагментом в хранилище.

9. Способ по п. 1, в котором генерирование по меньшей мере одного модифицированного аудиофрагмента выполняется сервером в потоковом режиме, и при этом способ дополнительно содержит этап, на котором передают сервером видеофрагмент видеофайла с по меньшей мере одним модифицированным аудиофрагментом на пользовательское устройство.

10. Способ по п. 9, дополнительно содержащий этапы, на которых:

получают сервером новый первый аудиофрагмент, представляющий дополнительное предложение на первом языке;

получают сервером новый второй аудиофрагмент, представляющий другое дополнительное предложение на втором языке, причем другое дополнительное предложение является переводом дополнительного предложения,

новый второй аудиофрагмент и подмножество вторых аудиофрагментов из последовательности вторых аудиофрагментов, формирующих новую последовательность вторых аудиофрагментов,

подмножество вторых аудиофрагментов, за исключением первого из последовательности вторых аудиофрагментов, использованных для генерирования по меньшей мере одной модифицированной части;

генерируют сервером новое множество возможных компоновок для новой последовательности вторых аудиофрагментов,

выбирают сервером новую целевую компоновку для новой последовательности вторых аудиофрагментов из нового множества возможных компоновок;

генерируют сервером новый модифицированный аудиофрагмент путем модификации первого из новой последовательности вторых аудиофрагментов в соответствии с новой целевой компоновкой; и

передают сервером другой фрагмент видеофайла с новым модифицированным аудиофрагментом на пользовательское устройство.

11. Сервер для генерирования модифицированных аудиоданных для видеофайла, причем видеофайл ассоциирован с аудиоданными, причем сервер сконфигурирован для:

получения последовательности первых аудиофрагментов,

данный фрагмент из последовательности первых аудиофрагментов представляет данное предложение на первом языке и ассоциирован с временной меткой в видеофайле;

получения последовательности вторых аудиофрагментов,

данный фрагмент из последовательности вторых аудиофрагментов представляет собой другое данное предложение на втором языке и ассоциирован с временной продолжительностью, при этом другое данное предложение является переводом данного предложения;

генерирования множества возможных компоновок для последовательности вторых аудиофрагментов,

данная возможная компоновка ассоциирована с возможными временными метками в видеофайле и возможными коэффициентами сжатия для соответствующих из последовательности вторых аудиофрагментов;

выбора целевой компоновки для последовательности вторых аудиофрагментов из множества возможных компоновок,

при этом сервер сконфигурирован для выбора на основе целевой компоновки, имеющей минимальную оценку потери из числа оценок потери, ассоциированных с соответствующими из множества возможных компоновок, при этом данная оценка потери для данной возможной компоновки генерируется на основе:

(i) различий между временными метками, ассоциированными с соответствующими из последовательности первых аудиофрагментов и соответствующими из данной возможной компоновки, и

(ii) наличия перекрытия между данной парой вторых аудиофрагментов, если: вторые аудиофрагменты из последовательности вторых аудиофрагментов скомпонованы в видеофайле в соответствии с возможными временными метками и возможными коэффициентами сжатия; и

генерирования по меньшей мере одного модифицированного аудиофрагмента для видеофайла в качестве перевода аудиоданных с использованием целевой компоновки.

12. Сервер по п. 11, в котором данный фрагмент из последовательности первых

аудиофрагментов представляет собой звуковой сигнал формы волны.

13. Сервер по п. 11, в котором данный фрагмент из последовательности первых аудиофрагментов представляет собой мел-спектрограмму.

14. Сервер по п. 11, который дополнительно сконфигурирован для генерирования последовательности вторых аудиофрагментов на основе последовательности первых аудиофрагментов.

15. Сервер по п. 11, в котором временная метка, ассоциированная с данным фрагментом из последовательности первых аудиофрагментов, является временной меткой начала, а данный фрагмент из последовательности первых аудиофрагментов дополнительно ассоциирован с временной меткой окончания в видеофайле, при этом возможные временные метки данной возможной компоновки, включающие в себя возможные временные метки начала и возможные временные метки окончания для соответствующих вторых аудиофрагментов, при этом данная оценка потери дополнительно генерируется на основе:

комбинации различий между начальными временными метками, ассоциированными с соответствующими из последовательности первых аудиофрагментов, и соответствующими возможными временными метками начала из данной возможной компоновки,

комбинации различий между временными метками окончания, ассоциированными с соответствующими из последовательности первых аудиофрагментов, и соответствующими временными метками окончания из данной возможной компоновки.

16. Сервер по п. 11, в котором временная метка, ассоциированная с данным из последовательности первых аудиофрагментов, является центральной временной меткой, при этом возможные временные метки данной возможной компоновки включают в себя возможные центральные временные метки для соответствующих вторых аудиофрагментов, при этом данная оценка потери дополнительно генерируется на основе:

комбинации различий между центральными временными метками, ассоциированными с соответствующими из последовательности первых аудиофрагментов, и соответствующими возможными центральными временными метками из данной возможной компоновки.

17. Сервер по п. 11, в котором оценка потери дополнительно генерируется на основе комбинации возможных коэффициентов сжатия для данной возможной компоновки.

18. Сервер по п. 11, который сконфигурирован для генерирования по меньшей мере одного модифицированного аудиофрагмента в автономном режиме и дополнительно сконфигурирован для хранения видеофайла с по меньшей мере одним измененным аудиофрагментом в хранилище.

19. Сервер по п. 11, который сконфигурирован для генерирования по меньшей мере одного модифицированного аудиофрагмента в потоковом режиме и дополнительно сконфигурирован для передачи видеофрагмента видеофайла с по меньшей мере одним модифицированным аудиофрагментом на пользовательское устройство.

20. Сервер по п. 19, который дополнительно выполнен с возможностью:
получения нового первого аудиофрагмента, представляющего дополнительное предложение на первом языке;

получения нового второго аудиофрагмента, представляющего другое дополнительное предложение на втором языке, причем другое дополнительное предложение является переводом дополнительного предложения,

новый второй аудиофрагмент и подмножество вторых аудиофрагментов из последовательности вторых аудиофрагментов, формирующих новую последовательность вторых аудиофрагментов,

подмножество вторых аудиофрагментов, за исключением первого из последовательности вторых аудиофрагментов, использованных для создания по меньшей мере одной модифицированной части;

генерирования нового множества возможных компоновок для новой последовательности вторых аудиофрагментов,

выбора новой целевой компоновки для новой последовательности вторых аудиофрагментов из нового множества возможных компоновок;

генерирование нового модифицированного аудиофрагмента путем модификации первого из новой последовательности вторых аудиофрагментов в соответствии с новой целевой компоновкой; и

передачу другого фрагмента видеофайла с новым модифицированным аудиофрагментом на пользовательское устройство.