

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ(12) **ЗАЯВКА НА ИЗОБРЕТЕНИЕ**

(21)(22) Заявка: 2017108533, 25.09.2015

Приоритет(ы):

(30) Конвенционный приоритет:
25.09.2014 US 14/496,538(43) Дата публикации заявки: 17.09.2018 Бюл. №
26(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 15.03.2017(86) Заявка РСТ:
US 2015/052194 (25.09.2015)(87) Публикация заявки РСТ:
WO 2016/049439 (31.03.2016)

Адрес для переписки:

129090, Москва, ул. Б.Спасская, 25, строение 3,
ООО "Юридическая фирма Городисский и
Партнеры"

(71) Заявитель(и):

**МАЙКРОСОФТ ТЕКНОЛОДЖИ
ЛАЙСЕНСИНГ, ЭлЭлСи (US)**

(72) Автор(ы):

**ПРОКОФЬЕВА Анна (US),
ДЖЕЛЫКЙИЛМАЗ Фетхие Асли (US),
ХАККАНИ-ТУР Дилек З. (US),
ХЕК Ларри (US),
СЛАНИ Малкольм (US)**(54) **ВЗГЛЯД ДЛЯ ПОНИМАНИЯ РАЗГОВОРНЫХ ЯЗЫКОВ В МНОГОМОДАЛЬНЫХ
ДИАЛОГОВЫХ ВЗАИМОДЕЙСТВИЯХ**

(57) Формула изобретения

1. Машинореализованный способ для понимания и разрешения ссылок на визуальные элементы в визуальном контексте диалоговой вычислительной системы, при этом способ содержит этапы, на которых:

- идентифицируют визуальные элементы, доступные для пользовательского взаимодействия в визуальном контексте;

- принимают, из одного или более устройств ввода, соединенных с диалоговой вычислительной системой, пользовательский ввод, ассоциированный с одним или более визуальными элементами в визуальном контексте, причем пользовательский ввод содержит:

- произнесение, извлекаемое из речевого ввода, ссылающегося на конкретный визуальный элемент из одного или более визуальных элементов; и

- ввод с помощью взгляда, ассоциированный, по меньшей мере, с некоторыми из одного или более визуальных элементов, причем, по меньшей мере, некоторые из одного или более визуальных элементов включают в себя конкретный визуальный элемент;

- извлекают лексические признаки и признаки взгляда, по меньшей мере, частично на основе визуальных элементов и пользовательского ввода; и

- определяют конкретный визуальный элемент, по меньшей мере, частично на основе

лексических признаков и признаков взгляда.

2. Машинореализованный способ по п. 1, в котором визуальный контекст представляет собой веб-обозреватель в свободной форме или прикладной интерфейс.

3. Машинореализованный способ по п. 1, в котором ввод с помощью взгляда содержит ввод с помощью взгляда, ассоциированный, по меньшей мере, с намеченным визуальным элементом, или ввод с помощью положения головы, ассоциированный, по меньшей мере, с намеченным элементом, при этом ввод с помощью положения головы пользователя служит в качестве прокси для ввода с помощью взгляда.

4. Машинореализованный способ по п. 1, дополнительно содержащий этап, на котором вычисляют вероятности, ассоциированные с отдельными визуальными элементами из визуальных элементов, чтобы определять конкретный визуальный элемент, причем вероятности, по меньшей мере, частично основаны на лексических признаках и признаках взгляда.

5. Машинореализованный способ по п. 1, в котором извлечение признаков взгляда содержит этапы, на которых:

- идентифицируют множество точек фиксации, ассоциированных с вводом с помощью взгляда;

- группируют предварительно определенное число из множества точек фиксации между собой в кластер; и

- идентифицируют центроид кластера в качестве конкретной точки фиксации для извлечения признаков взгляда.

6. Машинореализованный способ по п. 1, в котором конкретный визуальный элемент ассоциирован с действием, и способ дополнительно содержит, по меньшей мере, частично на основе идентификации конкретного визуального элемента, этап, на котором инструктируют выполнение действия, ассоциированного с намеченным визуальным элементом в визуальном контексте.

7. Машиночитаемый носитель, имеющий машиноисполняемые инструкции, которые при выполнении конфигурируют компьютер с возможностью осуществлять способ по любому из пп. 1-6.

8. Диалоговая вычислительная система для понимания и разрешения ссылок на визуальные элементы в визуальном контексте, причем диалоговая вычислительная система содержит:

- устройство речевого ввода;

- устройство слежения;

- машиночитаемые носители;

- один или более процессоров; и

- один или более модулей на машиночитаемых носителях, выполняемых посредством одного или более процессоров, причем один или более модулей включают в себя:

- приемный модуль, сконфигурированный с возможностью принимать:

- произнесение пользователя, транскрибированное из речевого ввода, принимаемого из устройства речевого ввода, причем речевой ввод ссылается на конкретный визуальный элемент из множества визуальных элементов, представленных в пользовательском интерфейсе, ассоциированном с визуальным контекстом; и

- ввод с помощью взгляда, принятый из устройства слежения, причем ввод с помощью взгляда направлен на один или более из множества визуальных элементов, представленных в пользовательском интерфейсе, ассоциированном с визуальным контекстом;

- модуль извлечения, сконфигурированный с возможностью извлекать набор признаков, по меньшей мере, частично на основе множества визуальных элементов, произнесения пользователя и ввода с помощью взгляда; и

- модуль анализа, сконфигурированный с возможностью идентифицировать конкретный визуальный элемент, по меньшей мере, частично на основе набора признаков.

9. Диалоговая вычислительная система по п. 8, дополнительно содержащая модуль отображения, сконфигурированный с возможностью отображать множество визуальных элементов в пользовательском интерфейсе.

10. Диалоговая вычислительная система по п. 8, в которой модуль извлечения дополнительно сконфигурирован с возможностью извлекать признаки тепловой карты, по меньшей мере, частично на основе ввода с помощью взгляда и множества визуальных элементов.

11. Диалоговая вычислительная система по любому из пп. 8-10, в которой набор признаков включает в себя, по меньшей мере:

- лексические признаки, при этом лексические признаки представляют лексическое сходство между текстом, ассоциированным с отдельными визуальными элементами из множества визуальных элементов, и произнесением пользователя; и

- признаки взгляда, при этом признаки взгляда представляют расстояния между ограничивающими прямоугольниками, ассоциированными с отдельными визуальными элементами, и точками фиксации, ассоциированными с вводом с помощью взгляда, в предварительно определенные времена.

12. Диалоговая вычислительная система по п. 11, в которой модуль анализа дополнительно сконфигурирован с возможностью вычислять вероятности, ассоциированные с отдельными визуальными элементами из множества визуальных элементов, чтобы идентифицировать конкретный визуальный элемент, причем вероятности, по меньшей мере, частично основаны на лексических признаках и признаках взгляда.

13. Диалоговая вычислительная система по п. 12, в которой модуль анализа дополнительно сконфигурирован с возможностью идентифицировать конкретный визуальный элемент, по меньшей мере, частично на основе конкретного элемента, имеющего наибольшую вероятность из всех вычисленных вероятностей, ассоциированных с множеством визуальных элементов.

14. Диалоговая вычислительная система по п. 12, в которой модуль анализа дополнительно сконфигурирован с возможностью:

- классифицировать лексические признаки в первом процессе;

- классифицировать признаки взгляда во втором процессе, причем второй процесс выполняется во время, отличающееся от первого процесса; и

- по меньшей мере, частично на основе классификации лексических признаков и классификации признаков взгляда:

- вычислять вероятности, ассоциированные с отдельными визуальными элементами из множества визуальных элементов, чтобы идентифицировать конкретный визуальный элемент; и

- идентифицировать конкретный визуальный элемент, по меньшей мере, частично на основе вычисленных вероятностей.