

REIVINDICACIONES

1. Un método para identificar la ocurrencia de una variación semántica de una frase en un pasaje por al menos un procesador, el método comprende:

basado en una representación textual de la frase, calcular un vector de incrustación de frases, que representa un significado semántico de la frase;

10 obtener una representación textual del pasaje, que comprende una pluralidad de palabras;

extraer, de la representación textual del pasaje, al menos un conjunto jerárquico de secuencias anidadas de palabras, en donde cada secuencia de un conjunto jerárquico forma un subconjunto de palabras de una secuencia posterior en el conjunto jerárquico;

20 para cada secuencia, calcular un vector de incrustación de secuencias correspondiente, que representa un significado semántico de la secuencia;

para uno o más vectores de incrustación de secuencias, calcular un valor de similitud vectorial correspondiente, que representa la

25

HOJA DE SUSTITUCIÓN (REGLA 26)

similitud con el vector de incrustación de frases;

identificar una secuencia correspondiente a un valor máximo de similitud vectorial del uno o más valores de similitud vectorial; y determinar la secuencia identificada como una variación semántica de la frase, basado en el valor máximo de similitud vectorial.

2. El método de la reivindicación 1, en donde el al menos un conjunto jerárquico comprende una pluralidad de conjuntos jerárquicos.

3. El método de la reivindicación 2, en donde extraer la pluralidad de conjuntos jerárquicos comprende:

seleccionar una pluralidad de secuencias núcleo, cada una comprende una o más palabras del pasaje; y

para cada secuencia núcleo, producir un conjunto jerárquico respectivo de secuencias anidadas, cada una de las cuales comprende la secuencia núcleo y secuencias de palabras posteriores, que aumentan monótonamente.

4. El método de la reivindicación 3, en donde la representación textual del pasaje es una transcripción de una conversación, y en donde seleccionar las secuencias núcleo comprende:

seleccionar una sección de la transcripción,
asociada con un hablante específico en la
conversación; y
seleccionar las secuencias núcleo para
5 comprender una o más palabras de la sección
seleccionada.

5. El método de la reivindicación 3, en donde
seleccionar las secuencias núcleo comprende:

calcular, para una o más palabras del
10 pasaje, una o más etiquetas respectivas que
representan partes del discurso (POS); y
seleccionar las secuencias núcleo para
comprender una o más palabras del pasaje
basado en las etiquetas de POS calculadas.

15 6. El método de la reivindicación 3, en donde
seleccionar las secuencias núcleo comprende:

calcular, para una o más palabras del
pasaje, una o más métricas respectivas de
relevancia del término; y
20 seleccionar las secuencias núcleo para
comprender una o más palabras del pasaje
basado en las métricas calculadas de
relevancia del término.

7. El método de la reivindicación 1, en donde
25 calcular un vector de incrustación de secuencia de una
secuencia específica comprende:

obtener un modelo basado en aprendizaje automático (ML), preentrenado para mapear entre representaciones textuales de palabras y vectores de incrustación de palabras correspondientes;

inferir el modelo basado en ML en una o más palabras de la secuencia específica, para producir uno o más vectores de incrustación de palabras correspondientes, basado en dicho entrenamiento; y

calcular el vector de incrustación de secuencias como una función del uno o más vectores de incrustación de palabras.

8. El método de la reivindicación 7, en donde entrenar el modelo ML comprende:

inferir el modelo basado en ML en una primera palabra, para producir un primer vector de incrustación de palabras provisionales, que representa un significado semántico de la primera palabra;

inferir el modelo basado en ML en una segunda palabra, para producir un segundo vector de incrustación de palabras provisionales, que representa un significado semántico de la segunda palabra;

recibir un primer elemento de datos de anotación, que representa la similitud semántica entre la primera palabra y la segunda palabra;

5 calcular un valor de similitud vectorial, que representa la similitud entre los vectores de incrustación de palabras provisionales;

10 calcular un primer valor de función de pérdida, que representa una diferencia entre el valor de similitud vectorial y la similitud semántica, como se representa mediante el primer elemento de datos de anotación; y

15 entrenar el modelo ML para minimizar el primer valor de la función de pérdida.

9. El método de la reivindicación 7, en donde entrenar el modelo ML comprende:

20 inferir el modelo basado en ML en una primera secuencia, para producir un primer vector de incrustación de secuencias provisionales;

 inferir el modelo basado en ML en una segunda secuencia, para producir un segundo

25 vector de secuencia provisional;

recibir un segundo elemento de datos de anotación, que representa la similitud semántica de secuencias entre la primera secuencia y la segunda secuencia;

5 calcular un segundo valor de similitud vectorial, que representa la similitud entre los vectores de incrustación de secuencias provisionales;

 calcular un segundo valor de función de pérdida, que representa una diferencia entre

10 el segundo valor de similitud vectorial y la similitud semántica de secuencias, como se representa mediante el segundo elemento de datos de anotación; y

15 entrenar el modelo ML para minimizar el segundo valor de la función de pérdida.

10. El método de la reivindicación 7, en donde entrenar el modelo ML además comprende:

 recibir una representación textual de una

20 frase;

 recibir un elemento de datos de anotación de pasaje, que indica la existencia de una variación de la frase en el pasaje;

 inferir el modelo basado en ML en un

25 conjunto jerárquico de secuencias obtenidas a partir de dicho pasaje, para calcular un

HOJA DE SUSTITUCIÓN (REGLA 26)

valor máximo de similitud vectorial
provisional; y
entrenar el modelo ML de tal manera que un
valor del valor máximo de similitud
5 vectorial provisional corresponda a la
ocurrencia de una variación de la frase en
el pasaje, como se representa mediante la
anotación del pasaje.

11. El método de la reivindicación 10, en donde
10 el elemento de datos de anotación de pasaje está
desprovisto de información que indique la ubicación de la
variación de la frase dentro del pasaje recibido.

12. Un sistema para identificar la ocurrencia de
una variación semántica de una frase en un pasaje, el
15 sistema comprende: un dispositivo de memoria no
transitoria, en donde se almacenan módulos del código de
instrucciones, y al menos un procesador asociado con el
dispositivo de memoria, y configurado para ejecutar los
módulos del código de instrucciones, después de la
20 ejecución de dichos módulos del código de instrucciones,
el al menos un procesador se configura para:

basado en una representación textual de la
frase, calcular un vector de incrustación de
frases, que representa un significado
25 semántico de la frase;

obtener una representación textual del
pasaje, que comprende una pluralidad de
palabras;

5 extraer, de la representación textual del
pasaje, al menos un conjunto jerárquico de
secuencias anidadas de palabras, en donde
cada secuencia de un conjunto jerárquico
forma un subconjunto de palabras de una
secuencia posterior en el conjunto

10 jerárquico;

para cada secuencia, calcular un vector de
incrustación de secuencias correspondiente,
que representa un significado semántico de
la secuencia;

15 para uno o más vectores de incrustación de
secuencias, calcular un valor de similitud
vectorial correspondiente, que representa la
similitud con el vector de incrustación de
frases;

20 identificar una secuencia correspondiente a
un valor máximo de similitud vectorial del
uno o más valores de similitud vectorial; y
determinar la secuencia identificada como
una variación semántica de la frase, basado

25 en el valor máximo de similitud vectorial.

13. El sistema de la reivindicación 12, en donde el al menos un conjunto jerárquico comprende una pluralidad de conjuntos jerárquicos.

14. El sistema de la reivindicación 13, en donde
5 el al menos un procesador se configura además para extraer la pluralidad de conjuntos jerárquicos al:

seleccionar una pluralidad de secuencias núcleo, cada una comprende una o más palabras del pasaje; y
10 para cada secuencia núcleo, producir un conjunto jerárquico respectivo de secuencias anidadas, cada una de las cuales comprende la secuencia núcleo y secuencias de palabras posteriores, que aumentan monótonamente.

15. El método de la reivindicación 14, en donde la representación textual del pasaje es una transcripción de una conversación, y en donde el al menos un procesador se configura además para seleccionar las secuencias núcleo al:

20 seleccionar una sección de la transcripción, asociada con un hablante específico en la conversación; y
seleccionar las secuencias núcleo para comprender una o más palabras de la sección
25 seleccionada.

16. El sistema de la reivindicación 14, en donde el al menos un procesador se configura además para seleccionar las secuencias núcleo al:

5 calcular, para una o más palabras del pasaje, una o más etiquetas respectivas que representan partes del discurso (POS); y
seleccionar las secuencias núcleo para comprender una o más palabras del pasaje basado en las etiquetas de POS calculadas.

10 17. El sistema de la reivindicación 14, en donde el al menos un procesador se configura además para seleccionar las secuencias núcleo al:

15 calcular, para una o más palabras del pasaje, una o más métricas respectivas de relevancia del término; y
seleccionar las secuencias núcleo para comprender una o más palabras del pasaje basado en las métricas calculadas de relevancia del término.

20 18. El sistema de la reivindicación 12, en donde el al menos un procesador se configura además para calcular el vector de incrustación de secuencias de una secuencia específica al:

25 obtener un modelo basado en aprendizaje automático (ML), preentrenado para mapear entre representaciones textuales de palabras
HOJA DE SUSTITUCIÓN (REGLA 26)

y vectores de incrustación de palabras correspondientes;

inferir el modelo basado en ML en una o más palabras de la secuencia específica, para

5 producir uno o más vectores de incrustación de palabras correspondientes, basado en dicho entrenamiento; y

calcular el vector de incrustación de secuencias como una función del uno o más

10 vectores de incrustación de palabras.

19. El sistema de la reivindicación 18, en donde el al menos un procesador se configura además para entrenar el modelo ML adicionalmente al:

recibir una representación textual de una

15 frase;

recibir un elemento de datos de anotación de pasaje, que indica la existencia de una variación de la frase en el pasaje;

inferir el modelo basado en ML en un

20 conjunto jerárquico de secuencias obtenidas a partir de dicho pasaje, para calcular un valor máximo de similitud vectorial provisional; y

entrenar el modelo ML de tal manera que un

25 valor del valor máximo de similitud vectorial provisional corresponda a la

HOJA DE SUSTITUCIÓN (REGLA 26)

ocurrencia de una variación de la frase en el pasaje, como se representa mediante la anotación del pasaje.

20. Un método para identificar la ocurrencia de una variación semántica de una frase en un pasaje por al menos un procesador, el método comprende:

basado en una representación textual de la frase, calcular un vector de incrustación de frases, que representa un significado semántico de la frase;

obtener una representación textual del pasaje, que comprende una pluralidad de n-gramas,

extraer, a partir de la representación textual del pasaje, una pluralidad de secuencias de n-gramas;

para cada secuencia, calcular un vector de incrustación de secuencias correspondiente, que representa un significado semántico de la secuencia;

para uno o más vectores de incrustación de secuencias, calcular una similitud con el vector de incrustación de frases; y

determinar la secuencia identificada como una variación semántica de la frase, con base en la similitud calculada.

HOJA DE SUSTITUCIÓN (REGLA 26)