

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 al:

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

20 para cada secuencia núcleo, generar un conjunto jerárquico respectivo de secuencias anidadas al anexar iterativamente una o más palabras que son adyacentes a la secuencia núcleo en el pasaje, en donde:

5 cada secuencia anidada en el
conjunto jerárquico respectivo
comprende la secuencia núcleo y al
menos una palabra adicional que
precede directamente y/o sigue
directamente la secuencia núcleo
dentro del pasaje y forma un
subconjunto de palabras de una
secuencia posterior en el conjunto
10 jerárquico respectivo; y
las secuencias anidadas se ordenan
al aumentar la longitud de tal
manera que cada secuencia posterior
incluye toda la secuencia anterior y
15 al menos una palabra adyacente
adicional;
para cada secuencia anidada, calcular un
vector de incrustación de secuencias
correspondiente, que representa un
20 significado semántico de la secuencia
anidada;
para uno o más vectores de incrustación de
secuencias, calcular un valor de similitud
vectorial correspondiente, que representa la
25 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 1, 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 comprender una o más palabras de la sección seleccionada.

4. El método de la reivindicación 1, en donde seleccionar las secuencias núcleo comprende:

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.

5. El método de la reivindicación 1, en donde
5 seleccionar las secuencias núcleo comprende:

calcular, para una o más palabras del pasaje, una o más métricas respectivas de relevancia del término; y

10 seleccionar las secuencias núcleo para comprender una o más palabras del pasaje basado en las métricas calculadas de relevancia del término.

6. El método de la reivindicación 1, en donde
calcular un vector de incrustación de secuencia de una
15 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;
20

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
25 dicho entrenamiento; y

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

7. El método de la reivindicación 6, en donde
5 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
10 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
15 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;

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

calcular un primer valor de función de pérdida, que representa una diferencia entre
25 el valor de similitud vectorial y la

similitud semántica, como se representa mediante el primer elemento de datos de anotación; y

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

8. El método de la reivindicación 6, en donde entrenar el modelo ML comprende:

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

10

 inferir el modelo basado en ML en una segunda secuencia, para producir un segundo vector de secuencia provisional;

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

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

20

 calcular un segundo valor de función de pérdida, que representa una diferencia entre el segundo valor de similitud vectorial y la similitud semántica de secuencias, como se

25

representa mediante el segundo elemento de
datos de anotación; y
entrenar el modelo ML para minimizar el
segundo valor de la función de pérdida.

5 9. El método de la reivindicación 6, en donde
entrenar el modelo ML además comprende:

recibir una representación textual de una
frase;

10 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
conjunto jerárquico de secuencias obtenidas
a partir de dicho pasaje, para calcular un
15 valor máximo de similitud vectorial
provisional; y

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

10. El método de la reivindicación 9, en donde
el elemento de datos de anotación de pasaje está
25 desprovisto de información que indique la ubicación de la
variación de la frase dentro del pasaje recibido.

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

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

15 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 al:

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

25 para cada secuencia núcleo, generar un conjunto jerárquico respectivo de secuencias anidadas al anexar

iterativamente una o más palabras que son adyacentes a la secuencia núcleo en el pasaje, en donde:

- 5 cada secuencia anidada en el conjunto jerárquico respectivo comprende la secuencia núcleo y al menos una palabra adicional que precede directamente y/o sigue directamente la secuencia núcleo
- 10 dentro del pasaje y forma un subconjunto de palabras de una secuencia posterior en el conjunto jerárquico respectivo; y
- las secuencias anidadas se ordenan
- 15 al aumentar la longitud de tal manera que cada secuencia posterior incluye toda la secuencia anterior y al menos una palabra adyacente adicional;
- 20 para cada secuencia anidada, calcular un vector de incrustación de secuencias correspondiente, que representa un significado semántico de la secuencia anidada;
- 25 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;

5 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.

10 12. El sistema de la reivindicación 11, en donde el al menos un conjunto jerárquico comprende una pluralidad de conjuntos jerárquicos.

13. El método de la reivindicación 11, en donde la representación textual del pasaje es una transcripción
15 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 seleccionada.

14. El sistema de la reivindicación 11, en donde
25 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.

15. El sistema de la reivindicación 11, en donde
el al menos un procesador se configura además para
seleccionar las secuencias núcleo al:

10 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
15 basado en las métricas calculadas de
 relevancia del término.

16. El sistema de la reivindicación 11, en donde
el al menos un procesador se configura además para
calcular el vector de incrustación de secuencias de una
20 secuencia específica al:

 obtener un modelo basado en aprendizaje
 automático (ML), preentrenado para mapear
 entre representaciones textuales de palabras
 y vectores de incrustación de palabras
25 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
5 calcular el vector de incrustación de secuencias como una función del uno o más vectores de incrustación de palabras.

17. El sistema de la reivindicación 16, en donde
10 el al menos un procesador se configura además para entrenar el modelo ML adicionalmente al:

 recibir una representación textual de una frase;
 recibir un elemento de datos de anotación de
15 pasaje, que indica la existencia de una variación de la frase en el pasaje;
 inferir el modelo basado en ML en un conjunto jerárquico de secuencias obtenidas a partir de dicho pasaje, para calcular un
20 valor máximo de similitud vectorial provisional; y
 entrenar el modelo ML de tal manera que un valor del valor máximo de similitud vectorial provisional corresponda a la
25 ocurrencia de una variación de la frase en

el pasaje, como se representa mediante la anotación del pasaje.

18. 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 n-gramas,

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

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

20 para cada secuencia núcleo, generar secuencias respectivas de n-gramas añadiendo iterativamente una o más palabras que son adyacentes a la secuencia núcleo en el pasaje, en donde
25 cada secuencia de n-gramas comprende la secuencia núcleo y al menos una palabra

adicional que precede directamente y/o
sigue directamente la secuencia núcleo
dentro del pasaje;

5 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
10 vector de incrustación de frases; y

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