



Texto predictivo

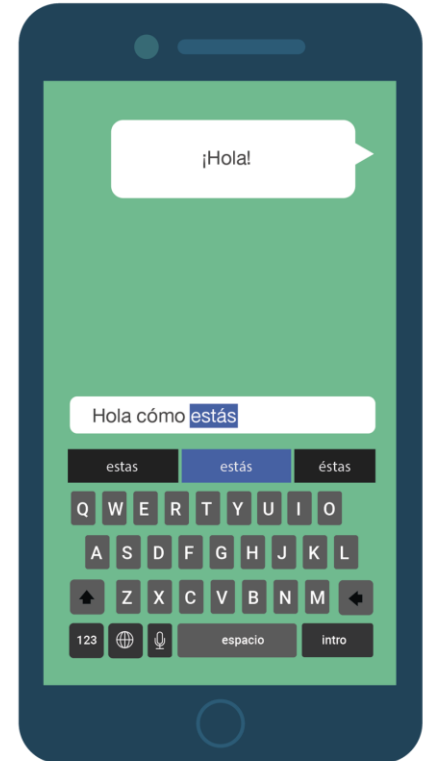


Texto predictivo



Texto predictivo

1. ¿Alguna vez han utilizado una aplicación que tenga texto predictivo? ¿Cuál es su opinión acerca de esta tecnología?
2. ¿Cómo piensan que esta aplicación es capaz de predecir las palabras que podríamos escribir?

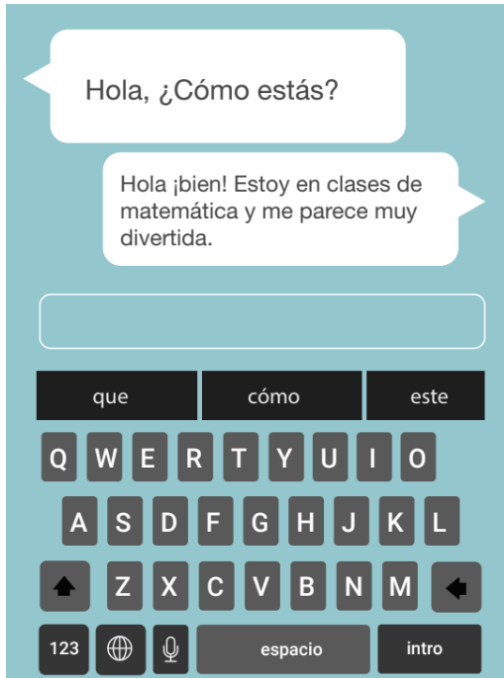


Texto predictivo

- Las aplicaciones de texto predictivo utilizan algoritmos que predicen cuál será la siguiente palabra que el usuario escribirá, considerando las palabras escritas previamente.
- Los algoritmos son secuencias de instrucciones que se utilizan para resolver un problema o llevar a cabo una tarea específica.



Texto predictivo



- En el caso de las aplicaciones de texto predictivo, los algoritmos se basan en probabilidades, estimadas a partir de los mensajes de texto previamente escritos por el usuario.

Presentación del problema:



Probabilidad condicional

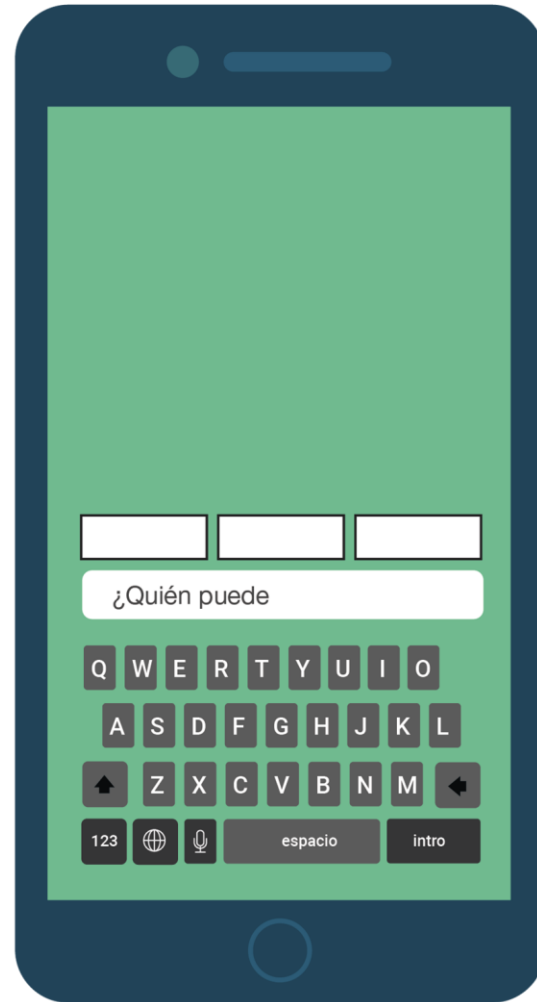
Sean A y B dos sucesos de un experimento aleatorio. La probabilidad de que ocurra A dado que ocurre B se representa por $P(A|B)$ y se puede determinar como:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \text{ siempre que } P(B) > 0$$

Es importante notar que en la probabilidad condicional el espacio muestral se reduce al evento B.

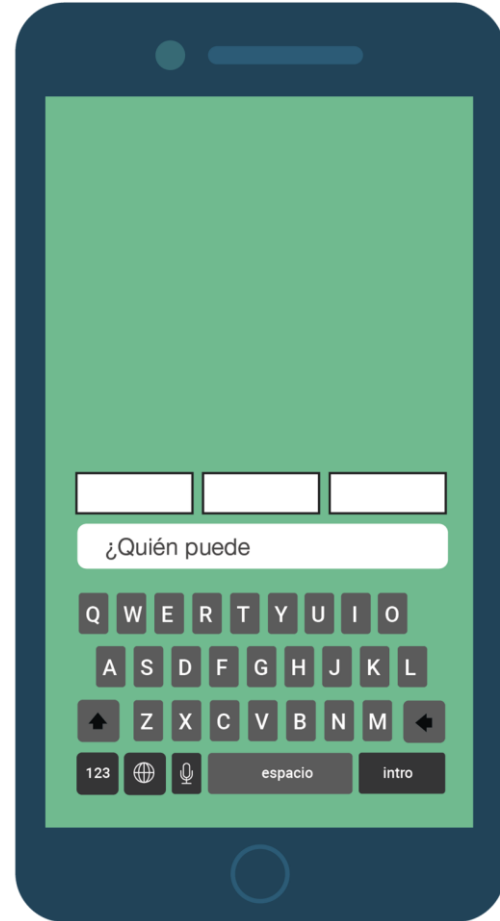
Actividad 1

1. A partir del texto ingresado, sugieran 3 palabras que el texto predictivo podría proponer.



Actividad 1

2. ¿Cuál fue el criterio utilizado para seleccionar las 3 palabras para la aplicación?
3. ¿Por qué creen que hubo tanta variación en las propuestas de palabras de los diferentes grupos?
4. ¿Cómo podemos identificar las palabras más probables de ser escritas después de la frase "¿Quién puede _____"?

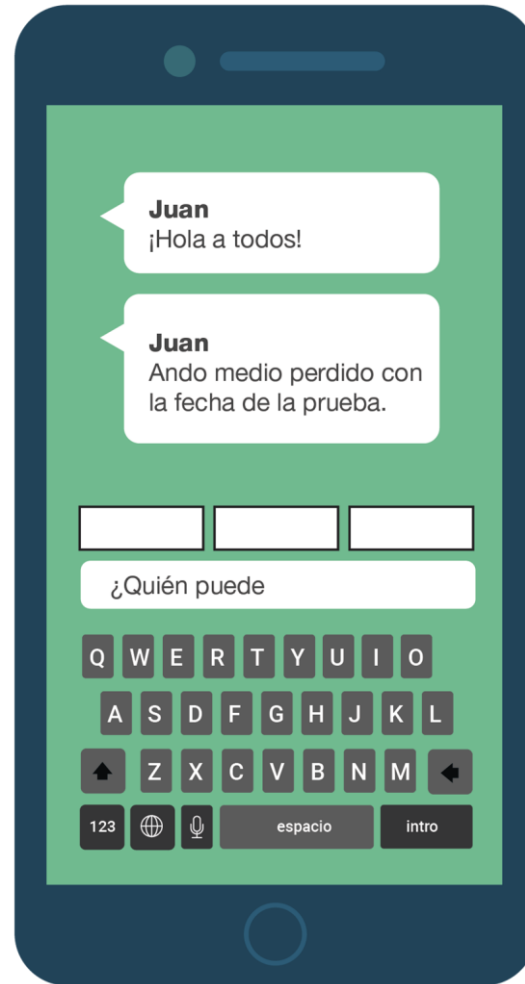


Comprendamos que:

2. El criterio consiste en seleccionar las palabras que son más comunes en el tipo de conversación que cada grupo se imaginó.
3. Hay variación en las palabras propuestas debido al tipo de conversación que cada grupo imaginó. Por ejemplo, si se piensa que es un chat de trabajo, las palabras seleccionadas van a diferir de las palabras elegidas al pensar en un chat entre amigos.
4. Para determinar cuáles son las palabras más probables que sigan a la frase dada, es necesario contar con más mensajes de texto del mismo tipo de conversación en el que se escribió esa frase.

Contexto:

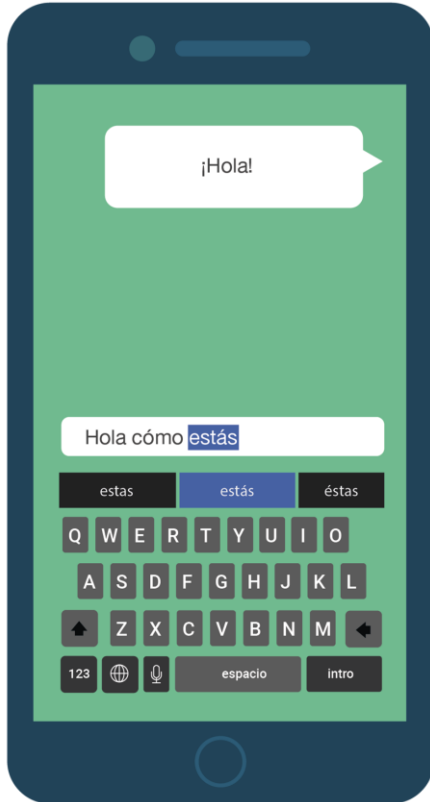
¿Cambiarían las palabras que proponen para el texto predictivo si conocieran los mensajes anteriores?



Reflexionemos:

- Los algoritmos de aprendizaje autónomo utilizados en las aplicaciones de texto predictivo imitan la forma en que los seres humanos aprenden y toman decisiones.
- Crear un algoritmo que considere el contexto completo de la conversación es difícil, por lo que se puede optar por uno más sencillo que solo tome en cuenta las palabras previas.

Notemos que:



- Sin datos previos, cualquier comparación de las probabilidades de posibles palabras se basan en creencias o experiencias personales.
- Para determinar de manera objetiva cuáles son las palabras más probables de continuar la frase, necesitamos contar con datos de mensajes de texto reales.

Actividad 2

La siguiente tabla contiene 20 mensajes de texto de un grupo de estudio:

¿Quién puede enviarme el archivo?	¿Quién puede decirme si hay tarea para mañana?	¿Quién puede preguntarle al profe por la tarea?	Gente, ¿quién puede prestarme su cuaderno para sacarle fotocopia?
Oye, ¿alguien puede decirme cómo se hace el ejercicio 4?	¿Quién puede enviarme el link para conectarse?	¿Quién puede ayudarme a traducir este texto en inglés?	¿Quién puede ayudarme con el trabajo de física?, no entiendo nada.
Kbros, quién puede prestarme una calculadora porfa.	¿Quién puede preguntarle al profe cuándo es la prueba?	Quién puede decirme cuál es la tarea para el viernes porfa.	¿Alguien quiere enviarme la materia por favor?
Chiquillos, ¿alguien sabe quién puede ayudarme con este ejercicio?	Gente, ¿quién puede ayudarme con la tarea de literatura? No sé por dónde empezar.	¿Quién puede enviarme la presentación que mandó la profe?	Oigan, ¿quién puede decirme las páginas del libro que hay que leer?
¿Quién puede hacer la portada del trabajo?	Chicos, ¿Pueden decirme qué entra en la prueba de historia?	Cabros, ¿quién puede ayudarme con el resumen del libro?	¿Alguien puede ayudarme con el trabajo de Química?

Actividad 2

1. Consideremos que A representa la palabra “enviarme” y B la frase “¿Quién puede”.
¿Cómo podemos calcular la probabilidad de que en un mensaje de texto aparezca la palabra A inmediatamente después que es escrita la frase B?

Guiémonos con algunas preguntas:

1. Sea $P(A)$ la probabilidad de que en un mensaje de texto aparezca la palabra A .
¿Cómo podemos calcular esta probabilidad?
1. ¿Cuál es el evento y el espacio muestral involucrados en la probabilidad $P(A)$?
1. ¿Es ésta la probabilidad que se pide determinar en esta actividad?

Respuestas:

- $P(A) = \frac{n(A)}{n(\Omega)} = \frac{4}{20} = 0,2$, donde $n(A)$ es el número de mensajes de textos que contienen A y $n(\Omega)$ es el número total de mensajes de textos considerados.
- El evento se refiere a la aparición de la palabra A en los 20 mensajes de texto, mientras que el espacio muestral está compuesto por el conjunto de esos 20 mensajes de texto.
- No, la probabilidad que se pide determinar es la probabilidad de que aparezca la palabra A inmediatamente después de la frase B, no la probabilidad de que aparezca la palabra A en general.

Consideremos lo siguiente:

- El evento se limita a los mensajes de texto que contienen la palabra A inmediatamente después de la frase B, y se excluyen los mensajes en que A no aparece justo después de la frase B.
- El espacio muestral se reduce a los mensajes de texto que contienen la frase B.

Consideremos lo siguiente:

Podemos calcular la probabilidad pedida como:

$$\frac{n_B(A)}{n(B)}$$

Donde $n_B(A)$ representa la frecuencia de la palabra A en el espacio muestral reducido a los mensajes de texto que contienen la frase B.



Contemos, para eso:

1. Marquen todas las palabras A (“enviarme”) que aparezcan en los mensajes de texto.
2. Tachen todas las oraciones que no contengan la frase B (“¿Quién puede”).
3. Cuenten todas las palabras A que quedaron sin tachar.



Contemos, para eso:

¿Quién puede <u>enviarme</u> el archivo?	¿Quién puede decirme si hay tarea para mañana?	¿Quién puede preguntarle al profe por la tarea?	Gente, ¿quién puede prestarme su cuaderno para sacarle fotocopia?
Oye, ¿alguien puede decirme cómo se hace el ejercicio 4?	¿Quién puede <u>enviarme</u> el link para conectarse?	¿Quién puede ayudarme a traducir este texto en inglés?	¿Quién puede ayudarme con el trabajo de física?, no entiendo nada.
Kbros, quién puede prestarme una calculadora porfa.	¿Quién puede preguntarle al profe cuándo es la prueba?	Quién puede decirme cuál es la tarea para el viernes porfa.	¿Alguien quiere <u>enviarme</u> la materia por favor?
Chiquillos, ¿alguien sabe quién puede ayudarme con este ejercicio?	Gente, ¿quién puede ayudarme con la tarea de literatura? No sé por dónde empezar.	¿Quién puede <u>enviarme</u> la presentación que mandó la profe?	Oigan, ¿quién puede decirme las páginas del libro que hay que leer?
¿Quién puede hacer la portada del trabajo?	Chicos, ¿Pueden decirme qué entra en la prueba de historia?	Cabros, ¿quién puede ayudarme con el resumen del libro?	¿Alguien puede ayudarme con el trabajo de Química?

$$n_B(A) \neq n(A) \quad n_B(A) = n(A \cap B)$$

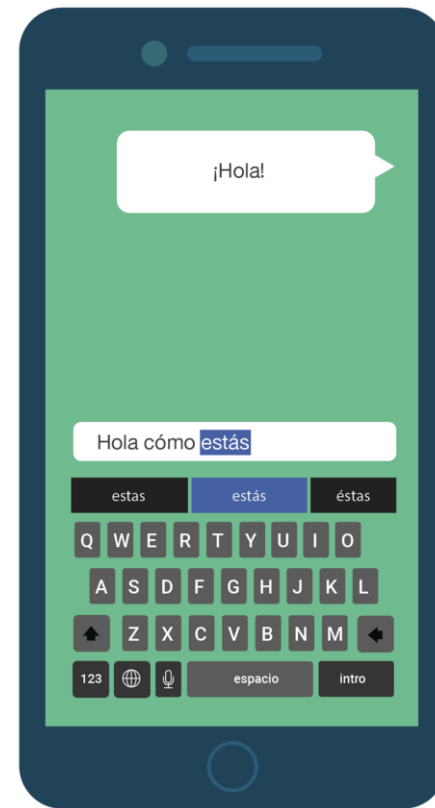
Consideremos lo siguiente:

La probabilidad solicitada se calcula como:

$$\frac{n(A \cap B)}{n(B)}$$

En donde:

- $n(A \cap B)$ es el número de mensajes de textos que contienen la frase B seguida de la palabra A
- $n(B)$ es el número de mensajes de texto que incluyen la frase B.



Luego:

La probabilidad solicitada se calcula como:

$$\frac{n(A \cap B)}{n(B)} = \frac{3}{16} = 0,1875$$



Actividad 2

La probabilidad condicional de que en un mensaje de texto aparezca la palabra A, dado que la frase previa es B. Esto se denota $P(A | B)$ y se puede calcular como:

$$P(A | B) = \frac{n(A \cap B)}{n(B)}$$

Notemos que:

Si amplificamos por

$$\begin{aligned} P(A | B) &= \frac{n(A \cap B)}{n(B)} \\ &= \frac{n(A \cap B) \cdot \frac{1}{n(\Omega)}}{n(B) \cdot \frac{1}{n(\Omega)}} \\ &= \frac{\frac{n(A \cap B)}{n(\Omega)}}{\frac{n(B)}{n(\Omega)}} \end{aligned}$$

Es decir, que la probabilidad condicional $P(A | B)$ es igual al cociente entre la probabilidad conjunta $P(A \cap B)$ y la probabilidad del evento que actúa como condición $P(B)$:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Actividad 2

2. A partir de los mensajes de texto entregados, considere las 6 posibles palabras A que aparecen inmediatamente después de la frase B = "¿Quién puede",

y

¿Quién puede enviarme cada archivo?	¿Quién puede decirme de la tarea para mañana?	¿Quién puede preguntarle a profe por la tarea?	Gente, ¿quién puede prestarme su cuaderno para sacarle fotocopia?
Oye, ¿alguien puede decirme cómo se hace el ejercicio 4?	¿Quién puede enviarme el link para conectarse?	¿Quién puede ayudarme a traducir este texto en inglés?	¿Quién puede ayudarme con el trabajo de física?, no entiendo nada.
Kbros, ¿quién puede prestarme una calculadora porfa.	¿Quién puede preguntarle al profe cuándo es la prueba?	¿Quién puede decirme cuál es la tarea para el viernes porfa.	¿Alguien quiere enviarme la materia por favor?
Chiquillos, ¿alguien sabe quién puede ayudarme con este ejercicio?	Gente, ¿quién puede ayudarme con la tarea de literatura? No sé por dónde empezar.	¿Quién puede enviarme la presentación que mandó la profe?	Oigan, ¿quién puede decirme las páginas del libro que hay que leer?
¿Quién puede hacer la portada del trabajo?	Chicos, ¿Pueden decirme qué entra en la prueba de historia?	Cabros, ¿quién puede ayudarme con el resumen del libro?	¿Alguien puede ayudarme con el trabajo de Química?

Actividad 2

2. A partir de los mensajes de texto entregados, considere las 6 posibles palabras A que aparecen inmediatamente después de la frase B = "¿Quién puede",

para cada una de ellas, calcule la probabilidad $P(A | B)$ y

$$\begin{aligned}
 P(\text{enviarme} | \text{¿Quién puede}) &= \frac{3}{16} = 0,1875 = 18,75\% \\
 P(\text{decirme} | \text{¿Quién puede}) &= \frac{3}{16} = 0,1875 = 18,75\% \\
 P(\text{prestarme} | \text{¿Quién puede}) &= \frac{2}{16} = 0,125 = 12,5\% \\
 P(\text{ayudarme} | \text{¿Quién puede}) &= \frac{5}{16} = 0,3125 = 31,25\% \\
 P(\text{preguntarle} | \text{¿Quién puede}) &= \frac{2}{16} = 0,125 = 12,5\% \\
 P(\text{hacer} | \text{¿Quién puede}) &= \frac{1}{16} = 0,0625 = 6,25\%
 \end{aligned}$$

Notemos que:

- Las probabilidades que hemos calculado son estimaciones basadas en las **frecuencias relativas** de los mensajes previos, es decir, son **probabilidades empíricas**.
- A medida en que aumente la cantidad de mensajes de texto que utilicemos en nuestros cálculos, las probabilidades empíricas se acercarán a sus respectivas **probabilidades teóricas**.

Problema:

Basándonos en las probabilidades estimadas a partir de los mensajes de texto disponibles, ¿cuáles son las tres palabras más apropiadas que una aplicación de texto predictivo debería sugerir a un usuario al escribir “¿Quién puede”?

Las tres palabras son:

- ayudarme (probabilidad de 31,25%)
- enviarme (probabilidad de 18,75%)
- decirme (probabilidad de 18,75%)

Problema:

Las tres palabras son:

- ayudarme (probabilidad de 31,25%)
- enviarme (probabilidad de 18,75%)
- decirme (probabilidad de 18,75%)

Describan el paso a paso que siguieron para sugerir las tres palabras, ayudarme, enviarme y decirme.

Actividad 3

Trabajen en grupo para completar un algoritmo (una secuencia de instrucciones paso a paso) que permita a una aplicación de texto predictivo sugerir las tres palabras más probables que el usuario desea escribir después de introducir una frase específica. Se proporcionan los dos primeros pasos del algoritmo:

Paso 1: Identifica la frase introducida por el usuario.

Paso 2: Selecciona todos los mensajes de texto del historial que contengan esa misma frase.

Paso 3:

Algoritmo básico de texto predictivo

1. Identifica la frase introducida por el usuario.
2. Selecciona todos los mensajes de texto del historial que contengan esa misma frase.
3. Identifica las palabras que siguen inmediatamente a la frase mencionada en el paso anterior.
4. Calcula la frecuencia de todas las palabras identificadas.
5. Determina la probabilidad de cada palabra identificada, dividiendo su frecuencia entre el número de mensajes de texto que contienen la frase del paso 1.
6. Identifica las tres palabras con mayor probabilidad.
7. Sugiere las tres palabras con mayor probabilidad al usuario.

Conclusiones

1. En esta actividad, desarrollamos un algoritmo básico de texto predictivo utilizando probabilidades condicionales.
2. Para ello, utilizamos datos de mensajes de texto y estimamos las probabilidades de que una determinada palabra A aparezca después de una frase B, utilizando las frecuencias relativas dadas por la expresión

$$\frac{n(A \cap B)}{n(B)}$$

Conclusiones

3. Cuanto mayor sea la cantidad de mensajes de texto que usemos, más se acercarán las probabilidades empíricas calculadas a las probabilidades teóricas correspondientes.
3. Las probabilidades involucradas en el problema son probabilidades condicionales $P(A|B)$, donde el espacio muestral se reduce al evento B.
3. Manipulando algebraicamente la expresión $\frac{n(A \cap B)}{n(B)}$, concluimos que:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



Texto predictivo

