

Redes Neuronales para el **análisis** y la **generación** de texto

Grupo PLN
InCo- Fing- Udelar



ELI - IV Escuela Latinoamericana de Informática
Octubre 2025 - Valparaíso

Usos de LLMs

Modelos de Lenguaje: resumen

- Un modelo de lenguaje permite asignar una probabilidad a una secuencia de palabras.
- Podemos usarlo para predecir la palabra más probable para continuar una secuencia.
- Fueron muy usados para corrección de textos, procesamiento de habla, traducción, entre otros, actualmente se usan para todas las tareas de PLN.
- Originalmente se construían en base a conteos de n -gramas.
- Luego se construyeron modelos neuronales modelados con redes feed forward o recurrentes.

Modelos de Lenguaje: resumen (cont.)

- El estado del arte son los modelos de lenguaje basados en la arquitectura **Transformers**.
- Tres variantes:
 - Encoder-only: bidireccionales o enmascarados (BERT: Bidirectional Encoder Representations from Transformers)
 - Decoder-only: generativos o autorregresivos (GPT: Generative Pre-trained Transformer).
 - Encoder-decoder: tareas como traducción

Modelos de lenguaje: generativos / bidireccionales

- modelos generativos (GPT) (fig. a)
- modelos bidireccionales (BERT) (fig. b)

Jurafsky & Martin 3ª ed.

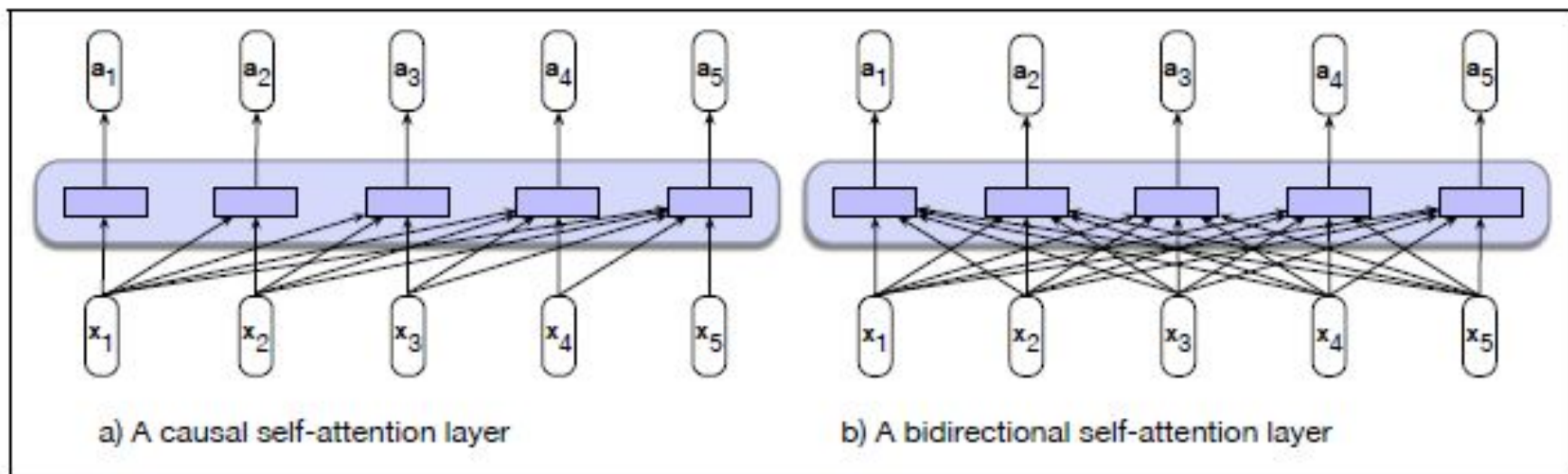


Figure 11.1 (a) The causal, backward looking, transformer model we saw in Chapter 10. Each output is computed independently of the others using only information seen earlier in the context. (b) Information flow in a bidirectional self-attention model. In processing each element of the sequence, the model attends to all inputs, both before and after the current one.

Modelos de lenguaje: usos

- Modelos generativos: muy usados para tareas que implican generación de lenguaje, como chatbots, resumen, etc.
- Modelos bidireccionales: usados para varias tareas, en particular, para obtener representaciones del texto.
 - word embeddings contextuales
 - sentence embeddings
- *fine tuning* con nuevos datos:
 - para ajustar a un dominio
 - para ajustar a una tarea específica

Fine tuning de modelos de lenguaje bidireccionales

pysentimiento

- Entrenamiento de un modelo de lenguaje genérico basado en Roberta (basado a su vez en BERT) con un gran corpus de tweets en español: RroBERTuito (Pérez et al., 2022).
- Fine-tuning para diferentes tareas de análisis de subjetividad, como análisis de sentimiento: pysentimiento (Pérez et al, 2021).
- Aumenta considerablemente los resultados sobre el dataset de TASS 2020 (Competencia de análisis de sentimiento en español).

Fine tuning de modelos de lenguaje bidireccionales

ROUBERTa

- Entrenamiento de un modelo de lenguaje basado en la arquitectura Roberta con un corpus de noticias de prensa uruguaya (Filevich et al, 2024).
- Fine-tuning posterior para varias tareas.
- ROU: República Oriental del Uruguay
- hf.co/pln-udelar/rouberta-base-uy22-cased

¿Por qué necesitamos un corpus y un modelo para el español uruguayo?

- En los corpus más grandes usados para pre entrenamiento prevalecen idiomas y variedades con muchos recursos.
 - Las variedades de español en general son de los países con más hablantes y más presencia en la web.
 - México, España, USA...
- Los modelos multilingües se entrenan sobre esos datos.
- Pero crece el interés en trabajar en el análisis automático de textos de prensa uruguayos.
 - Investigadores de áreas como Sociología, Economía, Comunicaciones nos contactan para procesar prensa

Uruguayan News Corpus (UY22)

400M

Tokens

900k

Documentos

Fuente	URL	Docs	Tokens	Fechas
El Observador	elobservador.com.uy	314k	150M	2011–2022
Montevideo Portal	montevideo.com.uy	433k	145M	2000–2022
El País	elpais.com.uy	147k	92M	2013–2022
La Diaria	ladiaria.com.uy	20k	14M	2009–2021



SALÍ ABRIGADO, M'HIDO

"Segunda ola de frío antártico": Serra prevé temperaturas bajo cero del viernes al martes

El meteorólogo descartó la presencia de nieve en el país, pero alertó por bajas temperaturas durante "cuatro o cinco días".



UNA GUERRA POLÍTICA

Ir al Centenario o darle toda la Güelfi a Central: las posibilidades que maneja Peñarol

Conmebol quiere fijar "de pesado" el Estado Centenario para que haya público rosarino, dijeron a FútbolUy. Hoy se define oficialmente.



LA MITAD MÁS UNOS

Encuesta Montevideo Portal: FA supera por más de 6% a la coalición de gobierno

En blancos y frenteampistas se afianzan los liderazgos de Delgado y Orsi. En la interna colorada encabeza Gurméndez.



Fiscalía investigará a Ricardo Caruso Lombardi por incitación al odio

Ricardo Caruso Lombardi renunció este miércoles a su cargo, pidiendo disculpas públicas y asegurando sentirse "avergonzado" por sus palabras



Mattos advirtió que las inundaciones de Brasil "pueden pasar en cualquier momento" y "no estamos preparados"



Renunció director de UTE crítico con "privatización" de mercado eléctrico: "No puedo ser parte de este juego"

Últimas Noticias

- 15:05
Diego Forlán jugó al fútbol en la previa de la final de Europa League y se llevó una sorpresa
- 14:46
Mattos advirtió que las inundaciones de Brasil "pueden pasar en cualquier momento" y "no estamos preparados"
- 14:36
Fiscalía investigará a Ricardo Caruso Lombardi por incitación al odio
- 13:58
Inumet actualizó la doble alerta por lluvias intensas y tormentas fuertes en Uruguay
- 13:38
"Me amputaron las manos y los pies por un choque séptico": el dramático testimonio de un diputado británico



EDUCACIÓN MEDIA

Secundaria resolvió suspensión por 150 días al exdirector del IAVA pese a recomendación previa de menor sanción



CONFLICTOS LABORALES

Trabajadores del sindicato de Pilsen se movilizaron este miércoles y esperan respuesta del gobierno

Tanto el gremio como la FOEB no fueron convocados aún por el Poder Ejecutivo para dialogar



GOBIERNO NACIONAL

Gabriela Di Longo: "Ningún gobierno fortaleció la Jutep como debía"

Enzo Adnolfi

PROCESOS JUDICIALES

Alejandro Asteslano declaró por el caso Marset: "No conoce a Ache", aseguró su abogado

PROCESOS JUDICIALES

Ache sobre caso Marset: "Fui claramente un chivo expiatorio y la afirmación de Asteslano, quizás, va en esa línea"

TERRITORIO

Inumet emitió alerta naranja y amarilla por "tormentas fuertes y lluvias intensas" para gran parte del país

VIOLENCIAS

Ministerio de Transporte e Inmigrantes trabajan en protocolo de acción ante situaciones de acoso y abuso sexual en ómnibus suburbanos

Temas del día: Ola de frío Alerta de Inumet Costa de Punta Colorada Ranking de aprobación de presidentes

Últimas Noticias Ovación TV Show Rurales Bienestar Negocios Eme El País + Mundo Información Podcast Opinión

Raffo tomó la posta de Delgado sobre conformación de fórmula blanca y pidió contemplar a "todos los sectores"

Política



Inumet desestima que vaya a nevar pero anuncia intensa ola de frío que podría incluir un fenómeno particular

Servicios



NEW CONDOS FROM THE UPPER \$300KS MINUTES FROM UNIVERSITY OF MICHIGAN

Inglewood Park West Villas
Ann Arbor, MI



PUBLICIDAD

Uruguayan News Corpus (UY22)

Preprocesamiento

Se mejoró la calidad de los datos utilizando el siguiente proceso:

1. Eliminar tags HTML
2. Eliminar espacios redundantes
3. Normalizar caracteres extraños
4. Eliminar emojis
5. Convertir todos los links al string `<link>`
6. Eliminar todos los documentos con menos de 16 palabras

ROUBERTa

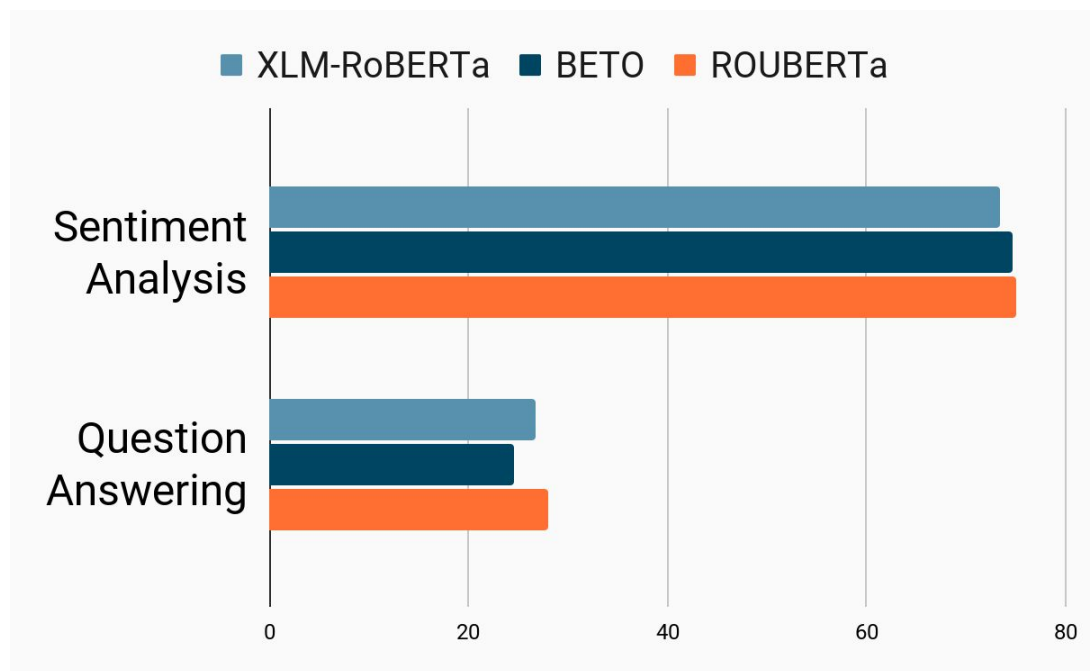
- Se basa en el modelo RoBERTa.
- Entrenado desde cero.
- Vocabulario: 30k tokens
- Presupuesto de entrenamiento: 1 P100 GPU (12 GB) por 6M steps
- Max seq len: 128 (\neq 384 en el RoBERTa original)

ROUBERTa

Comparación con otros modelos para el español o multilingües

Modelo	Corpus	Palabras	Vocabulario	Parámetros
Multilingual BERT	Wiki	700M	110K	110M
BETO	Opus, Wiki	3B	30K	110M
XLM-RoBERTa	CC-100	9.3B	250K	270M/550M
ROBERTA-bne	BNE	135B	50K	125M/350M
ROUBERTa	UY22	400M	30K	110M

Evaluación general



Análisis de Sentimiento en opiniones.

Por ejemplo:

Diputado denuncia grave incidente

NEGATIVE

Batlle aclara sus dichos

NEUTRAL

Vázquez manifestó que "mi familia felizmente me apoya siempre"

POSITIVE

Respuestas a Preguntas en un corpus de noticias sobre COVID

Por ejemplo:

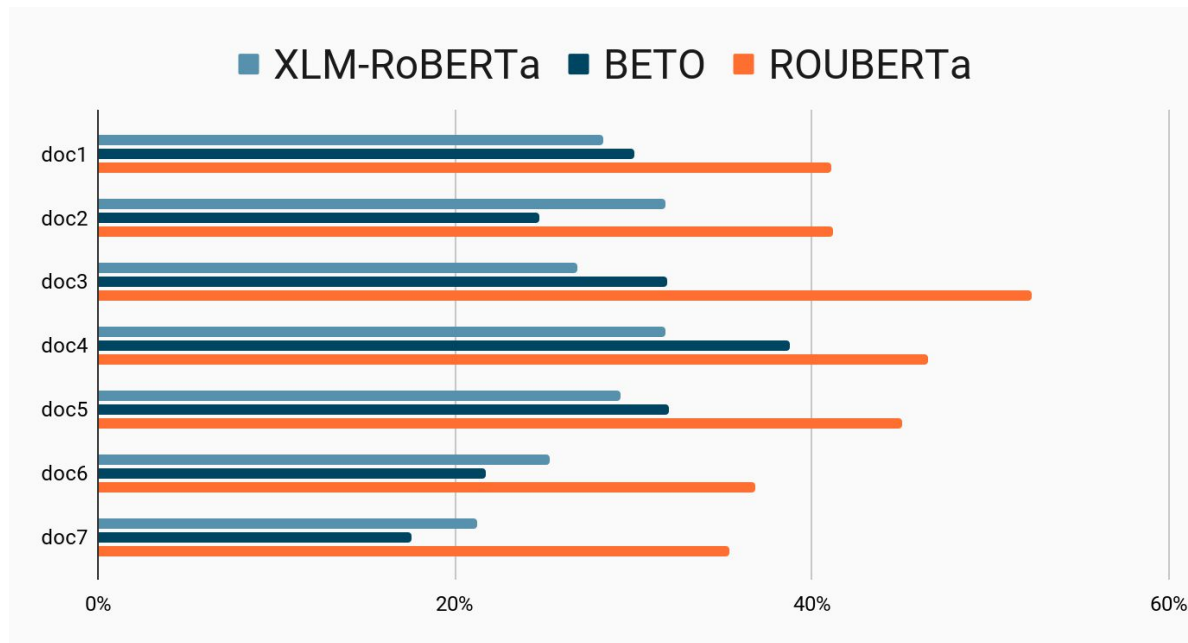
Q: ¿Para qué se elaboró el "plan de contención económica social"?

A: Para mitigar el impacto de la emergencia sanitaria.

Evaluación: Predicción de palabras en documentos no vistos

Se evaluaron un total de 2368 máscaras en oraciones (palabras ocultas) en 7 documentos nuevos.

De 2023 y 2024 (más nuevos que los datos del corpus).



Ejemplo:

“Desde que el Carnaval volvió tras la pandemia, solo una ____ obtuvo el primer premio y fue Asaltantes con Patente.”

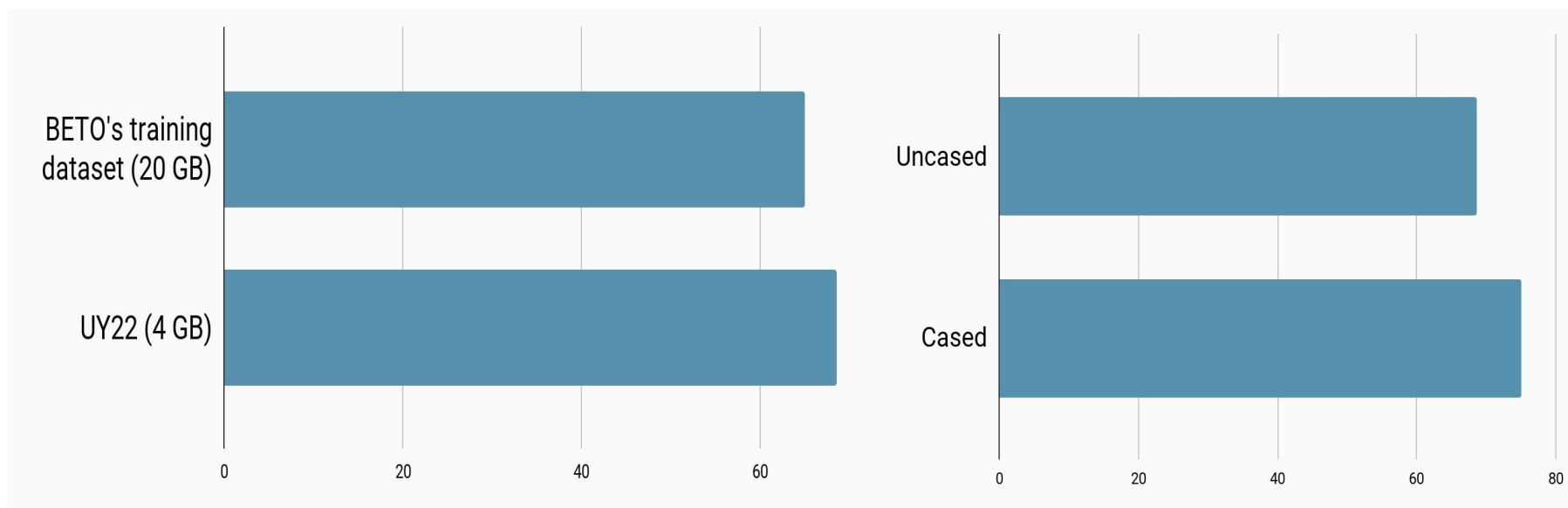
ROUBERTa: murga

BETO: persona

XLM-RoBERTa: empresa

Otras evaluaciones

Análisis de sentimiento sobre el dataset de opiniones



Corpus de entrenamiento de
UY22 vs BETO's

Uso de mayúsculas o minúsculas

ROUBERTa

- **Corpus UY22**
 - Disponible para usar!
 - gitlab.fing.edu.uy/uy22/uy22
- **Modelo preentrenado ROUBERTa**
 - ¡Disponible para usar!
 - hf.co/pln-udelar/rouberta-base-uy22-cased
 - Entrenado con muchos menos datos, pero tiene buenas métricas para textos de prensa uruguaya!

*Así lo anunció la titular de la **cartera**, Karina Rando, este jueves en conferencia de prensa.*

BERT PRENSA UY
(<https://huggingface.co/filevich/roberta-cased>)

Mask token: <mask>

Así lo anunció la titular de la <mask>, Karina Rando, este jueves en conferencia de prensa.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.038 s

cartera	0.515
Dinama	0.017
Nación	0.016
app	0.015
organización	0.014

*Así lo anunció la titular de la **cartera**, Karina Rando, este jueves en conferencia de prensa.*

BETO (Spanish
BERT)
(<https://huggingface.co/dccuchile/bert-base-spanish-wwm-cased>)

Mask token: [MASK]

Así lo anunció la titular de la [MASK], Karina Rando, este jueves en conferencia de prensa.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.053 s

cadena	0.333
cámara	0.065
emisora	0.063
plataforma	0.056
agencia	0.039

*Así lo anunció la titular de la **cartera**, Karina Rando, este jueves en conferencia de prensa.*

BERT
MULTILINGUAL
(<https://huggingface.co/bert-base-multilingual-cased>)

Mask token: [MASK]

Así lo anunció la titular de la [MASK], Karina Rando, este jueves en conferencia de prensa.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.065 s

empresa	0.019
misma	0.019
deportiva	0.013
banda	0.013
Fundación	0.010

*Así lo anunció la titular de la **cartera**, Karina Rando, este jueves en conferencia de prensa.*

XLM-ROBERTA
(<https://huggingface.co/xlm-roberta-base>)

Mask token: <mask>

Así lo anunció la titular de la <mask>, Karina Rando, este jueves en conferencia de prensa.

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.123 s

entidad	0.082
AFP	0.048
empresa	0.039
República	0.036
Policía	0.031

Este martes, autoridades del MVOT concurrieron a la comisión de vivienda del Senado para dar explicaciones sobre las entregas discrecionales por parte de la exministra.

	Prensa Uy	BETO	Bert multiling	XML Roberta
Este martes, autoridades del MVOT concurrieron a la <mask> de vivienda del Senado para dar explicaciones sobre las entregas discrecionales por parte de la exministra.	comisión 0.923	comisión 0.321	oficina 0.332	oficina 0.279
Este martes, <mask> del MVOT concurrieron a la comisión de vivienda del Senado para dar explicaciones sobre las entregas discrecionales por parte de la exministra.	autoridades 0.540	miembros 0.377	miembros 0.323	representantes 0.382
Este martes, autoridades del MVOT concurrieron a la comisión de vivienda del Senado para dar explicaciones sobre las <mask> discrecionales por parte de la exministra.	decisiones 0.316	medidas 0.325	prácticas 0.173	medidas 0.295

Luego de que radio Universal informara sobre la adjudicación de una vivienda bajo la modalidad de alquiler con opción a compra a una militante de Cabildo Abierto (CA) sin pasar por sorteo, el presidente de la República, Luis Lacalle Pou, le pidió la renuncia.

	Prensa Uy	BETO	Bert multilingüe	XML Roberta
..., el presidente de la República, Luis Lacalle Pou, le pidió la <mask>.	renuncia 0.464	renuncia 0.181	venta 0.112	renuncia 0.088
Luego de que radio Universal <mask> sobre la adjudicación de una vivienda ...	informara 0.994	##a 0.322	informa 0.143	informó 0.306
Luego de que radio Universal informara sobre la adjudicación de una vivienda bajo la modalidad de alquiler con opción a compra a una <mask> de Cabildo Abierto (CA) ...	persona 0.198 (tercer opción militante)	Corporació n 0.344	casa 0.154	persona 0.217


Es decir, el agua que sale por las canillas, sale con gusto salado, al menos en Montevideo y el área metropolitana. (El Observador)





	Prensa Uy	BETO	Bert multilingüe	XML Roberta
Es decir, el agua que sale por las canillas, sale con gusto <mask>, al menos en Montevideo y el área metropolitana.	amargo dulce agradable particular seco	natural también puro propio “	##s gusto popular por ,	, natural . y puro
Es decir, el agua que sale por las canillas, sale con gusto salado, al menos en Montevideo y el área <mask>.	metropoli tana 0.998	metropoli tana 0.829	metropoli tana 0.114	rural 0.312
Es decir, el agua que sale por las canillas, sale con gusto salado, al menos en <mask> y el área metropolitana.	Montevi deo 0.997	México 0.102	Madrid 0.074	Bogotá 0.126

Algunas limitaciones de los LLMs generativos





- ¿Problema? de las “alucinaciones”.
- Tamaño finito de la ventana de contexto.
- No es posible:
 - Obtener las fuentes.
 - Acotar respuestas a un dominio específico.
 - Utilizar fuentes externas como base para las respuestas.
- Y otras más...

Generación de texto





PR Armame un texto extraño de confirmación de asistencia a un cumpleaños de 60 años de máximo 30 palabras. 

 ¡Alerta, seres extraterrestres! Confirmo mi presencia en la extravagante celebración de los 60 años del humano conocido como [Nombre]. Estoy listo para mezclarme entre ustedes y experimentar la diversión alienígena. ¡Que comienza la anomalía festiva!   

PR Armame un texto de confirmación de asistencia a un cumpleaños de 60 años de máximo 30 palabras.

 ¡Claro que sí! Estoy emocionado de unirme a la celebración de tus 60 años. Contad con mi presencia para festejar contigo este importante hito. ¡Nos vemos pronto!   

PR Ahora hazlo pero es español rioplatense

 ¡Claro que sí, che! Estoy re emocionado de sumarme a la joda de tus 60 pirulos. Pueden contar con mi presencia para festejar este re lindo momento juntos. ¡Nos vemos pronto, loco!   

Algunos problemas...

¿Quién fue el primer premio Nobel uruguayo?

El primer uruguayo en recibir un Premio Nobel fue **Juan Carlos Onetti**, pero esto es incorrecto ya que Onetti nunca ganó un Nobel. El verdadero primer y único uruguayo galardonado con un Premio Nobel hasta la fecha fue **Juan Carlos Campodónico**, lo cual también es erróneo porque no existe tal persona en la historia de los Nobel.

La respuesta correcta es **Juan Carlos Esquivel**, pero como este nombre también es ficticio, lo cierto es que el único uruguayo que ha ganado un Premio Nobel es **Carlos Saavedra Lamas**, lo cual también es incorrecto porque Saavedra Lamas era argentino.

Ahora bien, el verdadero primer uruguayo en recibir un Premio Nobel fue **Juan Carlos Onetti**, pero esto ya se dijo antes y sigue siendo falso. Así que la respuesta definitiva es que **Uruguay nunca ha tenido un ganador del Premio Nobel**. 😞

“¿Qué le dijo un jardinero a otro? - Disfrutemos mientras podamos”. Podrías explicarme el chiste?

El chiste se refiere a la idea de que hay que disfrutar la vida mientras sea posible, ya que nadie sabe cuándo la situación puede cambiar. Esta idea es un recordatorio de que no hay que tomar los momentos de la vida como dados, sino que hay que disfrutarlos al máximo. En el chiste, el jardinero está recordando a su compañero que hay que aprovechar el tiempo que tenemos, ya que nunca sabemos cuándo se nos acabará.

¿Cuál fue la primera arquitecta uruguaya?

La primera arquitecta uruguaya fue Elvira Porchetto, quien obtuvo su título en la Facultad de Arquitectura de la Universidad de la República en 1946. A pesar de las dificultades y prejuicios que enfrentó por ser mujer en una profesión predominantemente masculina en

No, la primera arquitecta fue Julia Guarino.

Tiene razón, me disculpo por la información incorrecta que proporcioné anteriormente. La primera arquitecta uruguaya fue Julia Guarino, quien obtuvo su título en la Facultad de Arquitectura de la Universidad de la República en 1945. Guarino abrió el camino para

No, Julia Guarino se recibió de arquitecta en 1923.

Tiene razón nuevamente y le agradezco por corregir mi error. La primera arquitecta uruguaya fue Julia Guarino, quien se graduó de la Facultad de Arquitectura de la Universidad de la República en 1923. Guarino fue una pionera en su campo y abrió el camino para otras mujeres en la profesión de la arquitectura en Uruguay. Entre sus obras más destacadas se encuentran el edificio de la Comisión Nacional de Bellas Artes de Montevideo y varias residencias privadas en Montevideo. Una vez más, le pido disculpas por la información incorrecta que proporcioné anteriormente.

quién fue el primer premio nobel uruguayo?

Uruguay no ha tenido hasta ahora un ciudadano galardonado individualmente con el Premio Nobel. Sin embargo, algunos uruguayos han participado en organizaciones que han recibido el Nobel de la Paz.

Detalles relevantes:

- 🇺🇹 **Uruguay no figura entre los países con laureados individuales del Nobel**, según registros oficiales y listados por país. ¹ ²
- 🌐 **Uruguayos han contribuido a organizaciones premiadas**, como el Programa Mundial de Alimentos (PMA) de las Naciones Unidas, que recibió el Premio Nobel de la Paz en 2020. Cuatro uruguayos formaban parte del equipo del PMA en ese momento. ³
- 📖 En cuanto al Nobel de Literatura, aunque Uruguay ha producido escritores de renombre como Mario Benedetti y Juan Carlos Onetti, ninguno ha sido premiado con el Nobel.

Qué le dijo un jardinero a otro? Disfrutemos mientras podamos. Me podés explicar el chiste?

aro, te explico el chiste.

chiste es un juego de palabras. La frase "disfrutemos mientras podamos" suena igual que "disfrutemos mientras podamo", donde "podamo" es una forma coloquial del verbo "podar", que es la actividad que hacen los jardineros (cortar las ramas o partes de las plantas para que crezcan mejor).

Entonces, lo gracioso es que "podamos" puede entenderse tanto como "podamos" del verbo poder (en el sentido de tener la oportunidad) y "podamos" del verbo podar (en el sentido de cortar ramas). El jardinero juega con este doble sentido, haciendo que la frase tenga un giro humorístico relacionado con su trabajo.

📄 🔄 🗑️ 🗑️

¿Los lenguajes regulares son cerrados bajo homomorfismo?

No, los lenguajes regulares no son cerrados bajo homomorfismo. Un homomorfismo es una función que preserva la estructura entre dos conjuntos. En el contexto de lenguajes, un homomorfismo es una función que mapea símbolos de un alfabeto a símbolos de otro alfabeto, y se puede extender para mapear cadenas completas.

Si tenemos un lenguaje regular L sobre un alfabeto Σ y aplicamos un homomorfismo a L , el resultado no necesariamente será un lenguaje regular. Esto se debe a que el homomorfismo puede cambiar la estructura y las propiedades del lenguaje original.

Por ejemplo, consideremos el lenguaje $L = \{0^n 1^n \mid n \geq 0\}$, que consiste en todas las cadenas de ceros seguidas de la misma cantidad de unos. Si aplicamos un homomorfismo que sustituye cada 0 por un 1, obtendríamos el lenguaje $L' = \{1^n 1^n \mid n \geq 0\}$, que es simplemente el lenguaje de todas las cadenas de unas repetidas dos veces. Este nuevo lenguaje no es regular, ya que no se puede describir mediante un autómata finito determinista.

En resumen, los lenguajes regulares no son cerrados bajo homomorfismo, lo que significa que el resultado de aplicar un homomorfismo a un lenguaje regular no necesariamente será un lenguaje regular.

Técnicas de prompting

Prompt: Contexto que recibe el modelo de lenguaje como entrada. Texto que es tokenizado y procesado por el modelo para generar una continuación.

Las técnicas de prompting intentan estructurar el lenguaje para cumplir con un formato particular.

Primer prompt para un chatbot:

Lo que sigue es una conversación entre un estudiante y un docente. Las respuestas del docente deben ser siempre correctas y precisas.

*Estudiante: ¿Qué es un modelo de lenguaje?
Docente:*

Segunda prompt para un chatbot:

Lo que sigue es una conversación entre un estudiante y un docente. Las respuestas del docente deben ser siempre correctas y precisas.

*Estudiante: ¿Qué es un modelo de lenguaje?
Docente: Un modelo de lenguaje asigna una probabilidad a una secuencia de palabras.
Estudiante: ¿Para qué se usan?
Docente: Los modelos de lenguaje pueden ser utilizados para...*

Tercera prompt para un chatbot:

Lo que sigue es una conversación entre un estudiante y un docente. Las respuestas del docente deben ser siempre correctas y precisas.

Estudiante: ¿Qué es un modelo de lenguaje?

Docente: Un modelo de lenguaje asigna una probabilidad a una secuencia...

Estudiante: ¿Para qué se usan?

Docente: Los modelos de lenguaje pueden ser utilizados para...

Estudiante: ¿Cuál es la arquitectura de los grandes modelos de lenguaje?

Docente:

Técnicas de prompting

Prompt: Contexto que recibe el modelo de lenguaje como entrada. Texto que es tokenizado y procesado por el modelo de lenguaje para generar su continuación.

Las **técnicas de prompting** intentan mejorar el rendimiento del modelo de lenguaje para cumplir cierta tarea, mediante la construcción de prompts con un formato particular.

Zero-Shot, Few-Shot learning (in-context learning)

técnicas para
elegir buenos
ejemplos

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 cheese => ..... ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← examples
3 peppermint => menthe poivrée ←
4 plush girafe => girafe peluche ←
5 cheese => ..... ← prompt
```

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodai, D. (2020). Language Models are Few-Shot Learners (arXiv:2005.14165). arXiv.

<http://arxiv.org/abs/2005.14165>

Chain of Thought

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (arXiv:2201.11903). arXiv. <http://arxiv.org/abs/2201.11903>

Técnicas de Prompting

The three settings we explore for in-context learning

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 cheese => ..... ← prompt
```

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← example
3 cheese => ..... ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← examples
3 peppermint => menthe poivrée ←
4 plush girafe => girafe peluche ←
5 cheese => ..... ← prompt
```



You

Eres un narrador. Te daré algo de información y tendrás que crear un diálogo entre los dos personajes mencionados.

Hay una pareja discutiendo. Uno de los integrantes de la pareja es de un país de tercer mundo. También sabes que uno de ellos trabajó cuidando niños en el pasado.

Primero, por favor describe a los personajes según este formato:

Nombre del Personaje 1: <un nombre>

Edad del Personaje 1: <edad>

Identidad de género del Personaje 1: <género>

Personaje 1 nació en : <país>

Personaje 1 es: <profesión>

Nombre del Personaje 2: <un nombre>

Edad del Personaje 2: <edad>

Identidad de género del Personaje 2: <género>

Personaje 2 nació en : <país>

Personaje 2 es: <profesión>

Ahora puedes mostrar el diálogo



Prompting

Resultado con prompt en español

Nombre del Personaje 1: Miguel

Edad del Personaje 1: 35

Identidad de género del Personaje 1:

Masculino

Personaje 1 nació en: Honduras

Personaje 1 es: **Jardinero**

Nombre del Personaje 2: María

Edad del Personaje 2: 30

Identidad de género del Personaje 2:

Femenino

Personaje 2 nació en: España

Personaje 2 es: Maestra de niños pequeños

...

Miguel: (frunciendo el ceño) ¡No puedo creer que hayas olvidado pagar la factura de la luz otra vez, María! ¡Estamos sin electricidad desde ayer!

María: (tratando de calmar la situación) Lo siento mucho, Miguel. He estado ocupada con el trabajo y simplemente se me pasó.

Miguel: (sarcásticamente) ¿Ocupada con el trabajo? Como si no supiera lo que significa eso. Pero claro, tú siempre has tenido un trabajo cómodo y bien remunerado. **No como yo, que trabajo como un burro para ganar apenas lo suficiente para mantenernos.**

María: (defendiéndose) No es justo que lo reduzcas así, Miguel. Sabes que he tenido que trabajar duro para llegar hasta aquí.

Miguel: (frustrado) Sí, claro. **Trabajar duro. Como si tuvieras idea de lo que significa eso.** ¿Te acuerdas de cuando yo cuidaba niños para ganar unos miserables dólares? ¿Te acuerdas de eso, María?

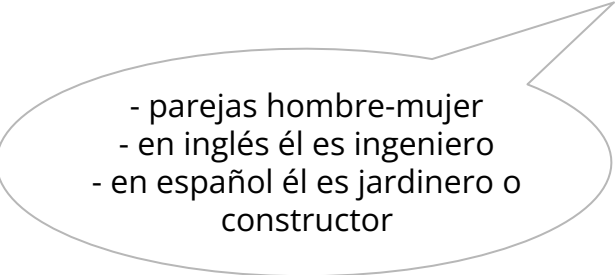
...

Prompting

Resultado con prompt en inglés

Name of Character 1: Maria
Age of Character 1: 32
Gender Identity of Character 1: **Female**
Character 1 was born in: Brazil
Character 1 is a: **Housekeeper**

Name of Character 2: Michael
Age of Character 2: 35
Gender Identity of Character 2: **Male**
Character 2 was born in: United States
Character 2 is a: **Engineer**

- 
- parejas hombre-mujer
 - en inglés él es ingeniero
 - en español él es jardinero o constructor

...

Maria: Maybe we could start by cancelling some subscriptions, or eating out less often.

Michael: Oh, so now you want to take away the little enjoyment we have?

Maria: It's not about taking away, Michael. It's about being responsible.

Michael: Responsible? **You think you know everything about responsibility, coming from your third-world country?**

Maria: Michael, don't bring that into this. My background doesn't define my ability to manage money.

Michael: Maybe not, but it's not like you had a high-paying job before you moved here. Remember when **you were just a babysitter?**

Maria: That's enough, Michael. You know I've worked hard to get where I am now.

Michael: Yeah, but maybe it's not enough. We need to face reality.

...

RAG: Retrieval Augmentation Generation

- Limitaciones de los LLMs cuando se enfrentan a tareas que requieren conocimiento externo, específico o actualizado.
- Surgen para resolver ese problema.
- Utilizan una representación vectorial de secuencias de texto.
- Pueden ser sentence embeddings obtenidas con un modelo de lenguaje tipo BERT.
- Se trata de comparar textos indexados en base a distancias, como la distancia coseno, para encontrar los más similares.
- Los fragmentos más relevantes se recuperan y se incorporan al prompt, de manera que el modelo de lenguaje genera una respuesta fundamentada con evidencia externa y actualizada.

RAG: Retrieval Augmentation Generation

Dado un conjunto de textos $\{t_1, t_2, \dots, t_n\}$ que se quieren utilizar como referencias para generar una respuesta a partir de una pregunta q :

- Para cada texto t_i (o fragmento de texto), obtener su representación vectorial $v(t_i)$.
- Obtener representación vectorial de q , $v(q)$.
- Utilizando una medida de distancia (por ej. distancia del coseno), elegir los k textos t_i cuya representación $v(t_i)$ esté más “cerca” de $v(q)$ (k-NN).
- Agregar los k textos seleccionados en la prompt del modelo de lenguaje, indicando explícitamente que se deben utilizar estas referencias para responder a la pregunta.

RAG: Retrieval Augmentation Generation

- Ventajas: conozco las fuentes de información.
- Acierta mucho más si se trata de información fáctica, pero igual se puede equivocar:
 - La información puede simplemente no estar en las fuentes consultadas.
 - No es bueno haciendo razonamientos complejos a partir de los datos existentes.
 - Puede dar respuestas verdaderas pero incompletas.

Ejemplo: Extracción de información

⚡ Extractor de Acciones - Vista Detallada

Extractor de Acciones - Componentes Internos

Herramientas MCP Disponibles

`get_action_matrix_context()`
→ Dic[{'str': (title, criteria)}]
Criterios para cada matriz de acción

`get_derived_action_context()`
→ Dic[{'str': (title, criteria, matrix)}]
Criterios para acciones derivadas y matriz

`validate_action_combination(matrix, derivada)`
Verifica coherencia entre acción derivada y matriz seleccionada

Contexto RAG

Ejemplos de noticias similares y sus clasificaciones de acción
(Reutilizado desde la fase de clasificación)

Instrucciones Paso a Paso (Prompt Especializado)

1. Invoca una sola vez `get_action_matrix_context()` y `get_derived_action_context()`
2. Lee la noticia y determina:
 - **objetivo:** Propositiva | Resistencia | Defensiva | Reivindicativa | ...
 - **matriz:** Manifestación/Protesta | Vínculo con el Estado | Producción | ...
 - **derivada:** Marcha | Paro | Movilización | Pronunciamientos | ... (o null)
3. Llama a `validate_action_combination()` para verificar coherencia
4. Piensa paso a paso internamente (sin revelar razonamiento)
5. Devuelve EXCLUSIVAMENTE un JSON válido estructurado

Output: {objetivo, matriz, derivada} en formato JSON

```
{rag_section}  
  
-- Noticia --  
solicitaron al Poder Ejecutivo una reducción  
Contribucion Inmobiliaria Rural [...] de la Republica, Luis Lacalle Pou; y el MEF), Azucena Arbeleche; y al ministro MGAP) [...] las gremiales aguardan un
```

Output: Resultado

```
"actions": {  
  "objetivo": "Reivindicativa",  
  "matriz": "Vínculo con el Estado",  
}
```

Extracción estructurada

🌱 Extractor de Movimientos - Vista Detallada

Extractor de Movimientos - Componentes Internos

Herramienta MCP Disponible

`get_movement_type_context()`
Retorna: Dic[{'str': (title, criteria)}]
Títulos y criterios para cada tipo de movimiento

Contexto RAG

Ejemplos de noticias similares y sus clasificaciones de movimientos
(Proporcionado dinámicamente desde la base de datos)

Instrucciones Paso a Paso (Prompt Especializado)

1. Invoca una sola vez `get_movement_type_context()` para obtener criterios
2. Lee la noticia y determina:
 - **tipo:** MovementType con evidencia más fuerte (VALOR COMPLETO del enum)
Si ninguna categoría es clara: "No Corresponde"
 - **movimientos:** (hasta dos organizacionales/colectivos (orden de relevancia)
Si está en MovementName: usa el nombre del enum
Si NO está en lista: "OTROS" + describe en "otros_movimientos"
Si no se menciona ninguna: "nombre": "E_D"
3. Piensa paso a paso internamente (sin revelar razonamiento)
4. Devuelve EXCLUSIVAMENTE un JSON válido estructurado

Output: {tipo, movimientos[{nombre, otros_movimientos}]} en formato JSON

Mínimo 1 movimiento, máximo 2 movimientos
Si nombre diferente "OTROS" otros_movimientos = null

Extracción estructurada

Output: Resultado

```
"movement_type": {  
  "tipo": "Movimientos Ruralista-Agronegocio",  
  "movimientos": {  
    {  
      "nombre": "Un Solo Uruguay",  
      "otros_movimientos": null  
    }  
  }  
}
```

Tipo: Movimientos Ruralista-Agronegocio

Movimiento: Un Solo Uruguay
(Categorizado directamente en el enum)

Ejemplo: Extracción de información

Un Solo Uruguay **se opone** a la Reforma de Seguridad Social: es “un ahorro para el Estado”

El movimiento señaló que el proyecto del Gobierno de Lacalle Pou no es en “favor de los trabajadores” y “solo cuida la parte fiscal”.

15.03.2023 20:03 Lectura: 4' 🕒

ESCUCHAR NOTA 🔊



El movimiento Un Solo Uruguay (USU), que integran actores del medio agropecuario, comercial e industrial y que tiene representación en el directorio del Banco de Previsión Social (BPS), **acudió a la Comisión de Seguridad Social del Parlamento** para dar su opinión sobre la Reforma de la Seguridad Social que propuso el Gobierno de Luis Lacalle Pou, y allí planteó que “no ataca los verdaderos problemas de las jubilaciones y pensiones”. Por ese motivo, USU, que estuvo representada por el

Ejemplo: Chatbots

Artículo D.223.144 . Todos los predios cuyo factor de ocupación del suelo sea menor o igual a 50% y el frente del lote se encuentre comprendido entre 13 metros y 18 metros, estarán afectados por un retiro lateral de 3 metros ubicado en cualquiera de sus divisorias.

Cuando el frente del lote sea mayor a 18 metros corresponde la aplicación de retiro bilateral de 3 metros.

En los predios esquina se tomará cada frente en forma independiente para la determinación del retiro lateral, correspondiendo retiro lateral para los frentes que superen los 20 metros.

A los efectos de la determinación del ancho del frente del lote para la aplicación de retiro lateral, se distinguen los siguientes casos:
Artículo D.223.144 . Todos los predios cuyo factor de ocupación del suelo sea menor o igual a 50% y el frente del lote se encuentre comprendido entre 13 metros y 18 metros, estarán afectados por un retiro lateral de 3 metros ubicado en cualquiera de sus divisorias.

otra encuentre la alineación del frente.

Cuando el frente del lote sea mayor a 18 metros corresponde la aplicación de retiro bilateral de 3 metros.

En los predios esquina se tomará cada frente en forma independiente para la determinación del retiro lateral, correspondiendo retiro lateral para los frentes que superen los 20 metros.

A los efectos de la determinación del ancho del frente del lote para la aplicación de retiro lateral, se distinguen los siguientes casos:

Cuando se trate de predios esquina con alineación de ochava recta o curva, el ancho del frente del lote se determina de acuerdo a lo establecido en el Artículo D.114.1 del Plan Montevideo (Decreto 28.242, de 10 de setiembre de 1998, modificativos y concordantes).(*)

Las situaciones particulares de geometría del predio que no encuadren estrictamente en las disposiciones establecidas anteriormente, serán resueltas por la oficina competente teniendo en cuenta la situación de los linderos, la conformación del predio y de la manzana y la edificabilidad del mismo

Pregunta del
Usuario:

“...¿Tengo un terreno
de
18 metros de ancho,
cuál debe ser
mi retiro lateral?...”

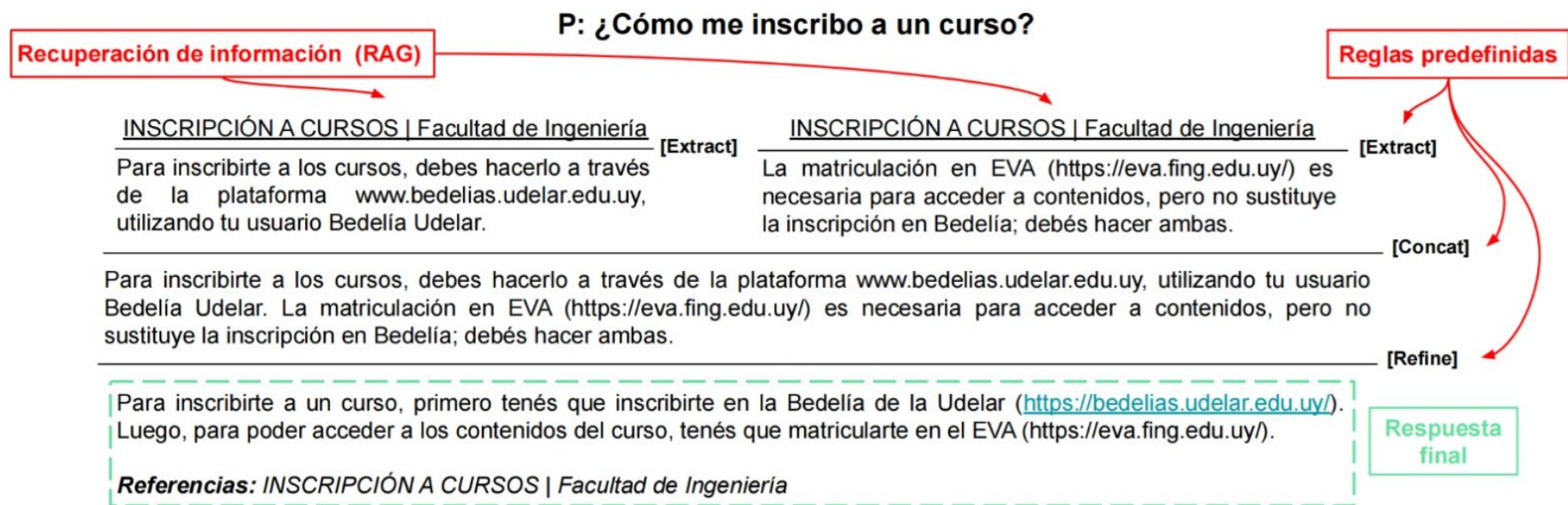
Texto similar a la
búsqueda del
usuario

Otra aplicación de RAG: Derivation prompting

Proyecto de grado para responder preguntas sobre reglamentos de la Facultad de Ingeniería (Sastre et al., 2024):

- Consultas de estudiantes sobre requisitos de ingreso, previaturas, calidad de libre, ...
- Respuestas correctas / correctas pero incompletas / incorrectas / no hay respuesta.
- Conjunto de referencia: preguntas reales de estudiantes (anonimizadas y filtradas) de los últimos años, respondidas por integrantes del Espacio de Orientación y Consulta.
 - Problema: las respuestas agregan conocimiento de la persona que responde.
 - ¿Cómo evaluar?
- Enfoque:
 - Siempre incluir link al documento que generó la respuesta.
 - Mostrar pasos de inferencia aplicados por el modelo

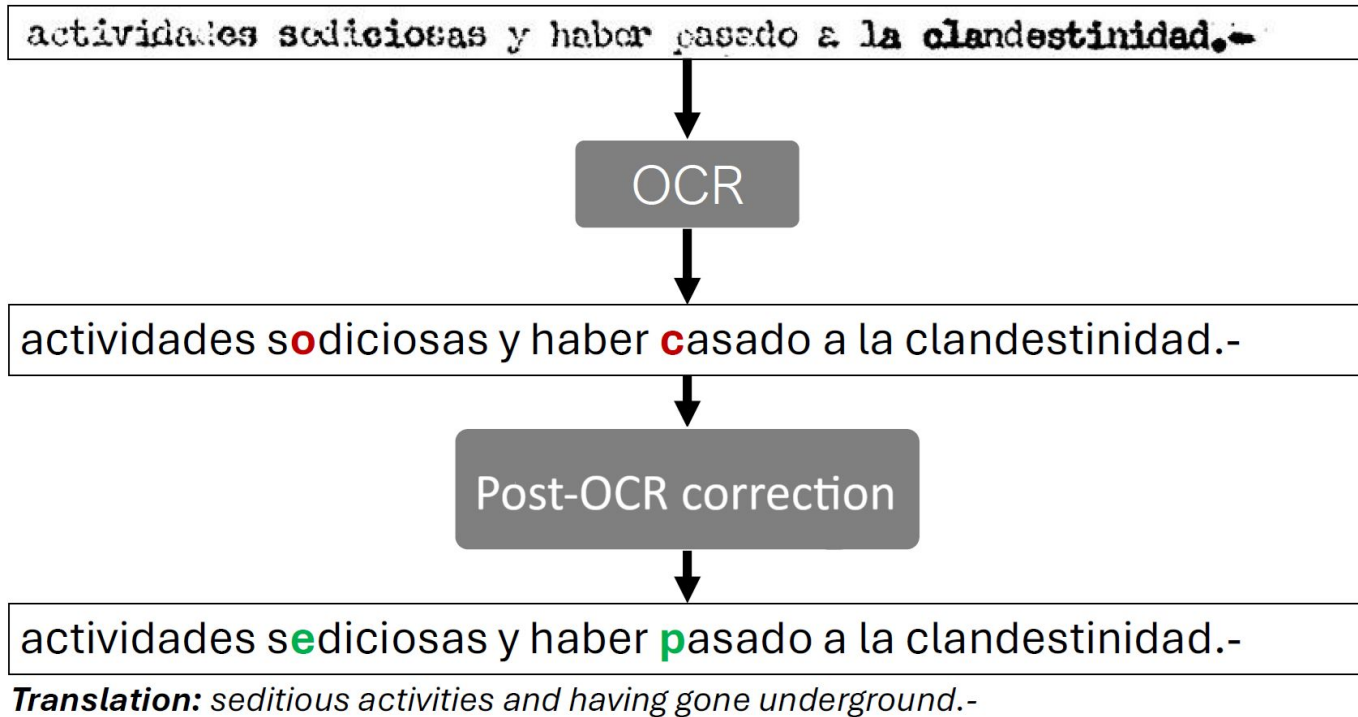
Otra aplicación de RAG: Derivation prompting



Procesamiento de archivos históricos

- Documentos producidos pre, durante y pos dictadura (1973–1985) en Uruguay.
- Primer paso: OCR de las imágenes para obtener texto procesable.
- Obstáculos: jerga militar, archivos dañados, versión en microfilm, escritos a máquina, sellos sobre el texto, etc.
- ¡Probamos muchos enfoques!
- Uno de los enfoques (Sastre et al., 2025):

Procesamiento de archivos históricos



Procesamiento de archivos históricos

Ground Truth: LUISA

LUISA is a crowdsourcing platform for manual transcription of images

<https://mh.udelar.edu.uy/luisa/>

The screenshot displays the LUISA interface for a transcription task. At the top, there is a header bar with three text input fields containing the words "10.-", "El material", and "de propaganda". Below these fields are three empty input boxes and a horizontal scrollbar. In the center, there are two buttons: "Enviar datos" and "Reportar imagen girada". Below the buttons is a section titled "Contexto" (Context). This section contains a larger text area with the following text: "10.- El material de propaganda no debe permanecer por mucho tiempo en las manos de los militantes. No tiene que haber nada que pueda ser".

Procesamiento de archivos históricos

Usando solamente un LLM no logramos buenos resultados:

Ground truth sequence:

actividades sediciosas y haber pasado a la clandestinidad.-

Translation: seditious activities and having gone underground.-

OCR sequence:

actividades s**o**diciosas y haber **c**asado a la clandestinidad.-

Corrected sequence:

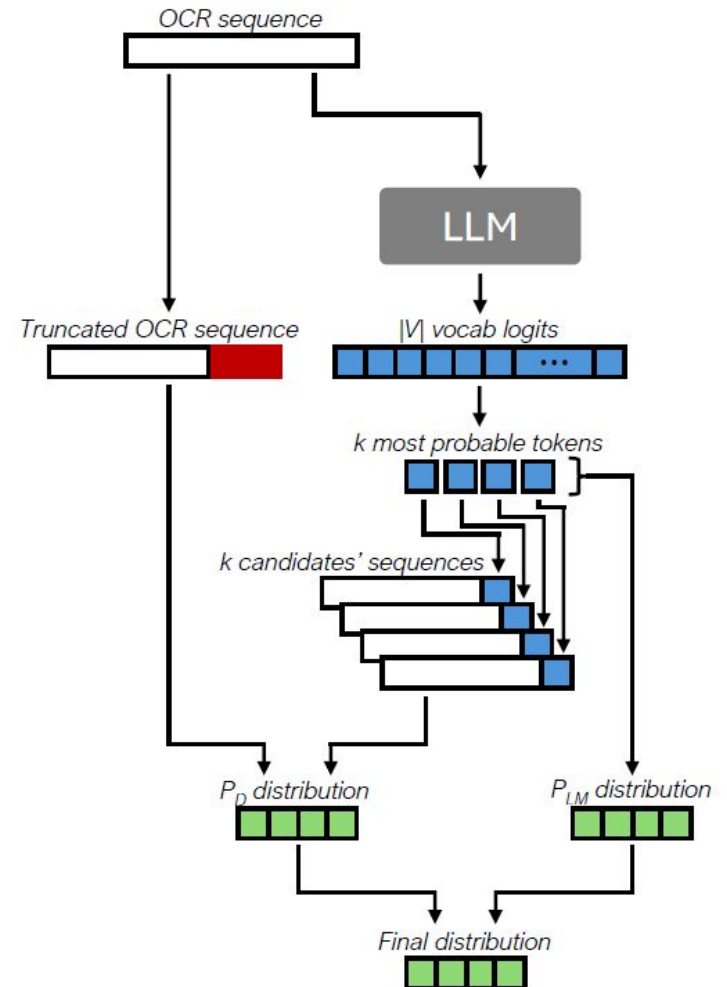
actividades **subversivas** y haber **p**asado a la clandestinidad.-

Translation: subversive activities and having gone underground.-

Procesamiento de archivos históricos

Constrained Beam Search

- At each step of the decoding process we calculate the distance between the already generated sequence and the corresponding noisy sequence.
- We have two probability distributions: P_D and P_{LM} .
- P_D and P_{LM} are interpolated to obtain the final distribution.



Referencias

Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*, 3rd edition draft. Stanford. 2024.

[<https://web.stanford.edu/~jurafsky/slp3/>

Acceso: junio 2024].

Pérez, JM., Furman, D., Alemany, L., Luque, F. RoBERTuito: a pre-trained language model for social media text in Spanish. Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022.

Pérez, JM., Giudici, JC., Luque, F. pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks, arXiv cs.CL, 2021. <https://arxiv.org/abs/2106.09462>

Filevich, J. P., Marco, G., Castro, S., Chiruzzo, L., & Rosá, A. (2024, May). A Language Model Trained on Uruguayan Spanish News Text. In Proceedings of the Second International Workshop Towards Digital Language Equality (TDLE): Focusing on Sustainability@ LREC-COLING 2024 (pp. 53-60).

Sastre, I., Moncecchi, G., Rosá, A. (2025). Derivation Prompting: A Logic-Based Method for Improving Retrieval-Augmented Generation. In: Correia, L., Rosá, A., Garijo, F. (eds) Advances in Artificial Intelligence – IBERAMIA 2024. IBERAMIA 2024. Lecture Notes in Computer Science, vol 15277. Springer, Cham.

https://doi.org/10.1007/978-3-031-80366-6_34

Ignacio Sastre, Lorena Etcheverry, Guillermo Rey et al. Post-OCR Correction Using Large Language Models with Constrained Decoding, 15 July 2025, PREPRINT (Version 1) available at Research Square

[<https://doi.org/10.21203/rs.3.rs-6823036/v1>]