

Recomendación de Textos Constitucionales: Un Enfoque Integrado de Similitud Jurídica en el Contexto de Procesos Constitucionales en Chile

Rodrigo Fuentes, José Domingo Silva, Lucas Valenzuela
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
{rfuente, jdsilvam, lucas.valenzuela}@uc.cl

Abstract

Este estudio aborda el desafío de recomendar textos constitucionales basándose en la similitud jurídica en el contexto de los recientes procesos constitucionales en Chile. Se presenta un sistema recomendador que combina, por un lado, una base de datos vectorial que recupera textos constitucionales según similitud semántica, y por otro lado, un modelo GPT afinado que filtra las recomendaciones considerando el parecido jurídico. Se compara este enfoque con un modelo de recomendación basado únicamente en la similitud semántica y se evalúa su desempeño utilizando métricas como MAP y nDCG. Los resultados indican que el sistema propuesto supera al modelo de referencia, destacando su capacidad para ofrecer recomendaciones más relevantes desde una perspectiva jurídica. Además, se exploran recomendaciones a nivel internacional, revelando patrones interesantes en la similitud jurídica entre los textos propuestos en los últimos años en Chile y constituciones de diferentes países y continentes, especialmente en América Latina. Este enfoque innovador contribuye al análisis y comprensión de similitudes jurídicas en textos constitucionales, brindando herramientas valiosas para contextos constitucionales dinámicos.

1. Introducción

Desde finales de 2019, se está debatiendo en Chile un cambio constitucional. En septiembre de 2022, los chilenos rechazaron la propuesta de Constitución redactada por la Convención Constitucional. Prontamente, las fuerzas políticas acordaron el inicio de un segundo proceso. Este se basó en dos instituciones principales, una Comisión de Expertos designada por los partidos políticos que redactó un anteproyecto, y un Consejo Constitucional democráticamente electo que modificó dicho texto para ofrecer una propuesta final. La nueva propuesta será plebiscitada el 17 de diciembre de 2023. En esta votación, los ciudadanos votarán si están en contra o a favor de la nueva propuesta de constitución. En este contexto, surge el interés de analizar el

contenido de estos textos constitucionales y crear un sistema que recomiende constituciones de otros países.

A la vez que Chile ha vivido este proceso político, también se han hechos grandes avances en el área de modelos de lenguaje, específicamente GPT, el cual se ha masificado desde el lanzamiento de ChatGPT y ha exhibido potencial en diversas tareas como, por ejemplo, recomendaciones y análisis de texto.

Uno de los desafíos más grandes que se pueden encontrar trabajando con textos legales y jurídicos es cómo poder calcular y medir la similitud entre dos textos. Un par de textos podrían ser muy similares en su estructura o el significado de las palabras que usan, pero pequeñas diferencias, como podrían ser una coma, o una palabra, podrían cambiar drásticamente el significado jurídico texto. Un claro ejemplo que se ha dado durante la discusión del Consejo Constitucional es el relacionado al derecho a la vida. “La ley protege la vida **de quien** está por nacer” y “La ley protege la vida **del que** está por nacer” en principio parecen dos oraciones idénticas, pero en el fondo una de ellas podría derogar la ley de interrupción del embarazo en tres causales. En este sentido, la recomendación basada en contenido se dificulta, ya que la similitud semántica de métodos de vectorización de texto, como TF-idf u otro tipo de embeddings, no necesariamente implica una similitud jurídica. La complejidad inherente de estos documentos y la creciente carga de información hacen imperativo el desarrollo de herramientas innovadoras que faciliten la identificación de textos relevantes.

Por la razón anterior, el objetivo de este trabajo será evaluar un modelo GPT que fue afinado sobre textos constitucionales con el objetivo de recomendar textos con similitud jurídica. Se contrastará con una recomendación pero basada únicamente en la similitud semántica, y por último, se compararán los resultados con un sistema recomendador basado en contenido más tradicional, que calcula las distancias coseno de los vectores TF-idf de los artículos de las constituciones. La motivación principal de este estudio es fomentar la participación ciudadana y el voto informado, a partir de la aplicación de GPT en el análisis documentos constitucionales para informar a los ciudadanos de manera clara y concisa.

2. Revisión bibliográfica

Debido a la novedad y carácter poco convencional de este trabajo, el estado del arte está más bien definido por las diferentes herramientas utilizadas como pasos intermedios en este sistema recomendador.

2.1. LLMs como sistemas recomendadores

El uso de *Large Language Models* (LLMs) como recomendadores es un área relativamente nueva, pero no por eso poco explorada.

Entre los trabajos que definen el estado del arte en el alineamiento de LLMs con sistemas recomendadores, se encuentra *TALLRec* [1], un *framework* que en búsqueda de accesibilidad, busca facilitar un proceso poco intensivo en *GPU*, a través de un *lightweight fine-tuning* del modelo. Otro trabajo en esta categoría es la proposición de *Chat-REC* [2], un paradigma que también busca ser accesible, mediante la transformación de datos históricos y perfiles de usuarios en *prompts* para presentarle a *Chat-GPT*.

2.2. Similitud entre textos jurídicos

La similitud entre textos es un campo ya muy explorado; desde los métodos para computarla hasta estudios comparativos entre los algoritmos existentes para ello [3] [4]. En este trabajo se computa con la distancia euclidiana entre *sentence embeddings*.

Este problema de similitud se complejiza en gran medida en el ámbito jurídico [5] [6]; área en que dos frases con una diferencia superficialmente pequeña, pueden tener implicancias legales muy grandes. Dado este hecho, es importante tomar medidas para darle relevancia a las particularidades del lenguaje que pueden pasar desapercibidas con métodos convencionales. Diversos estudios se han hecho acerca del como realizar determinar más precisamente la similitud entre textos legales, por ejemplo mediante el *fine tuning* de modelos NLP como BERT, GPT-2 y RoBERTa [7], o utilizando redes neuronales de manera más novedosa con *all-term cosine similarity* y *legal-term Cosine Similarity* [8]; ámbos métodos basados en las estadísticas *TF-IDF* de los términos presentes en los textos a comparar.

2.3. Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) introduce un enfoque novedoso para la generación de lenguaje en tareas intensivas en conocimiento [9]. El modelo RAG combina memoria paramétrica y no paramétrica al utilizar un recuperador para acceder a documentos relevantes y un generador para producir secuencias de salida basadas en la información recuperada. La principal contribución radica en la capacidad del modelo para incorporar tanto memoria preentrenada en forma de modelos seq2seq como acceso a una memoria no paramétrica, representada por un índice vectorial de Wikipedia. RAG demuestra un rendimiento

líder en tareas de preguntas y respuestas y generación de lenguaje, superando a modelos paramétricos y arquitecturas específicas para la recuperación y extracción de información. Además, se destaca la capacidad del modelo para generar lenguaje más específico, diverso y factual en comparación con enfoques paramétricos convencionales.

3. Metodología

Tal como muestra la Figura 1, la metodología se basa en cuatro componentes principales que serán explicados en detalle en cada una de las siguientes subsecciones.

3.1. Fine-Tuned Expert Agent

La existencia de una elevada similitud entre las representaciones vectoriales de dos textos no necesariamente denota una equivalencia jurídica. La tabla 1 ilustra la posibilidad de que dos pares de textos diferentes exhiban diferencias jurídicas sustanciales, a pesar de presentar similitudes vectoriales cercanas. Con el objetivo de abordar este problema, se plantea un *fine-tuning* de un modelo GPT con la capacidad de clasificar la similitud jurídica entre un par de textos, utilizando en el entrenamiento la base de datos que se detallará a continuación.

3.1.1. Base de datos de entrenamiento. La base de datos utilizada para el entrenamiento consta de pares de oraciones, los cuales fueron categorizados por un equipo de abogados constitucionalistas pertenecientes al Instituto Res Publica. Este conjunto de datos fue diseñado específicamente en el contexto del Proceso Constitucional de Chile en el año 2023. Durante este proceso, un Consejo Constitucional, elegido democráticamente por la ciudadanía, generó una Propuesta de Nueva Constitución a partir de un anteproyecto redactado por una Comisión de Expertos designada proporcionalmente por los partidos políticos según la cantidad de escaños en ambas cámaras legislativas.

El Equipo Constitucional del Instituto Res Publica analizó un total de 807 pares de oraciones extraídas de ambos textos y las clasificó en cuatro categorías diferentes de relaciones:

- 1) **INTACTO (150 casos):** Este par de oraciones exhibe una similitud intacta, sin considerar diferencias irrelevantes como puntuación, formato o tildes.
- 2) **MENOR (388):** Se refiere a cambios en la redacción que no afectan el contenido de manera significativa o cuya implicancia jurídica se considerara de menor importancia.
- 3) **MAYOR (86):** Estos pares de oraciones abordan el mismo tema, pero presentan cambios en la redacción que afectan las implicancias jurídicas en un grado considerable.
- 4) **SUPRESIÓN (183):** En este caso, las dos oraciones no guardan relación alguna entre sí.

La figura 2 representa la distancia euclidiana entre los embeddings de los pares de datos en el conjunto de entrenamiento. Estos embeddings fueron generados mediante el

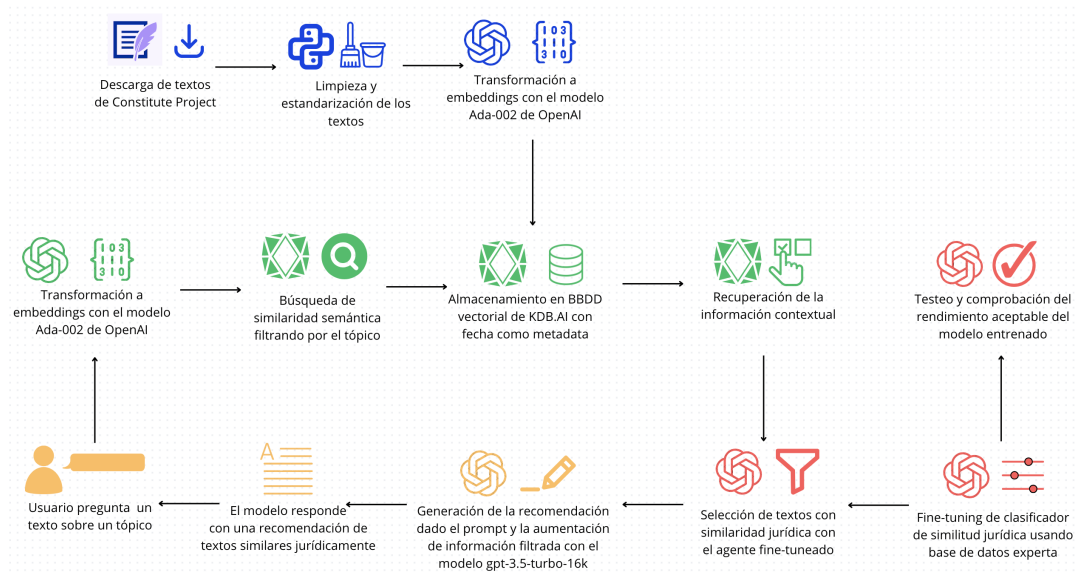


Figure 1. Diagrama de bloques

Relación	Similitud	Oraciones	Texto
Intacto	1,000	a) La familia es el núcleo fundamental de la sociedad	Comisión Experta
		b) La familia es el núcleo fundamental de la sociedad	Consejo Constitucional
Menor	0,653	a) El derecho al respeto y protección de su privacidad y la de su familia	Comisión Experta
		b) El derecho al respeto y protección de la vida privada y la de su familia	Consejo Constitucional
Mayor	0,658	a) La ley podrá regular, limitar o complementar el ejercicio de los derechos fundamentales	Comisión Experta
		b) Solo la ley podrá limitar o restringir el ejercicio de los derechos fundamentales	Consejo Constitucional
Supresión	0,644	a) El Fiscal de Asuntos Internos será nombrado por la Corte Suprema, a propuesta de una terna elaborada mediante un sistema de concurso público establecido en la ley institucional	Comisión Experta
		b) Los fiscales regionales serán nombrados por el Fiscal Nacional, a propuesta de una terna elaborada mediante un sistema de concurso público regulado en la ley institucional	Consejo Constitucional

TABLE 1. EJEMPLO DE PARES DE ORACIONES CON NIVELES PARECIDOS DE SIMILITUD PERO CON DIFERENTES IMPLICANCIAS JURÍDICAS

modelo GPT denominado ada-002. Aunque se observan tendencias que sugieren que, por ejemplo, pares de oraciones con modificaciones mínimas tienden a ser más similares que aquellos con cambios sustanciales, no se puede llegar a una conclusión jurídica basándose únicamente en esta observación.

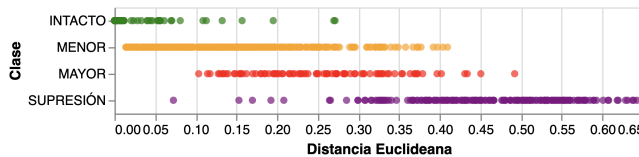


Figure 2. Similitudes de los pares de texto según etiquetas

3.1.2. Entrenamiento del *Fine-tuning*. Utilizando un conjunto de datos compuesto por 807 casos para el entrenamiento, se llevó a cabo el proceso de *fine-tuning* del modelo GPT 3.5-turbo-0613. Durante tres épocas, se entrenaron un total de 344,481 tokens, con un costo económico de \$2.76 dólares. La estructura de los datos siguió el siguiente formato, donde *sentence1*, *sentence2*, y *label* representan la oración del Anteproyecto, la propuesta de los

consejeros y la etiqueta otorgada por el equipo constitucional, respectivamente:

```
{ "messages":
  [{"role": "system", "content": "Eres un abogado experto constitucional que compara las diferencias entre pares de oraciones de constituciones diferentes"},
  {"role": "user",
    "content": f"a) {sentence1}
               b) {sentence2}"},
  {"role": "assistant",
    "content": {label}}]
}
```

3.1.3. Testeo de Fine-Tuned Expert Agent. Este agente está diseñado para clasificar pares de textos constitucionales de manera análoga a la evaluación realizada por un abogado experto. Se pone a prueba el modelo con otra base de datos del Instituto Res Publica que compara la Constitución Política de la República vigente con el Anteproyecto de la comisión Experta.

En total, se han evaluado 930 casos nuevos. Como se

muestra en las matrices de confusión de la figura 3, las etiquetas generadas por el modelo GPT tienden a presentar resultados positivos. Al complementar esto con las métricas de rendimiento de la tabla 2, se observa una alta precisión para las etiquetas INTACTO y MENOR, mientras que INTACTO y SUPRESIÓN muestran resultados destacados en *recall*. La etiqueta MAYOR tiende a obtener resultados deficientes, logrando predecir correctamente solo 10 de las 97 etiquetas. En total, la exactitud del modelo en el conjunto de prueba es del 76%, mientras que la precisión y el *recall* promedio ponderado son del 77% y 76%, respectivamente.

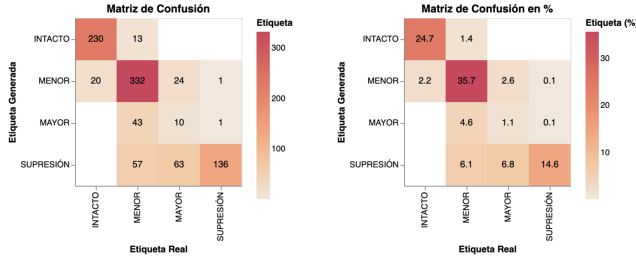


Figure 3. Matrices de confusión

Al analizar las confusiones de MAYOR con otras etiquetas, se observa que 63 de las 97 predicciones erróneas se clasifican como SUPRESIÓN, lo que representa un 65% de esta clase. Este hallazgo sugiere la existencia de un posible sesgo en el modelo en esa dirección. Este fenómeno podría atribuirse en parte a la baja frecuencia de las etiquetas MAYOR tanto en el conjunto de entrenamiento como en el de prueba.

Etiqueta	Precisión	Recall	F1-score	N
INTACTO	0.95	0.92	0.93	250
MENOR	0.88	0.75	0.81	445
MAYOR	0.19	0.10	0.13	97
SUPRESIÓN	0.53	0.99	0.69	138
Exactitud			0.76	930
Prom. macro	0.64	0.69	0.64	930
Prom. pond.	0.77	0.76	0.75	930

TABLE 2. MÉTRICAS DE RENDIMIENTO DEL CONJUNTO DE PRUEBA

3.1.4. Evaluación de etiquetas binarias. Al contrastar dos documentos, las etiquetas INTACTO o MENOR señalan niveles aceptables de similitud que permiten su comparación. En contraste, las etiquetas MAYOR y SUPRESIÓN indican textos con escasa similitud jurídica a pesar de poder tratar temas similares. Al combinar estas etiquetas en dos categorías, **ACEPTABLE** para INTACTO y MENOR, y **NO-ACEPTABLE** para MAYOR y SUPRESIÓN, las métricas de rendimiento aumentan en comparación con el caso original de cuatro etiquetas. La tabla 3 presenta los rendimientos de este *re-labeling*, donde la exactitud alcanza un 87% y la precisión a un 82%. Las principales equivocaciones provienen de predicciones erróneas entre etiquetas de tipo MENOR y MAYOR.

Etiqueta	Precisión	Recall	F1-score	N
ACEPTABLE	0.96	0.86	0.90	695
NO	0.68	0.89	0.77	235
Exactitud			0.87	930
Prom. macro	0.82	0.87	0.84	930
Prom. pond.	0.89	0.87	0.87	930

TABLE 3. MÉTRICAS DE RENDIMIENTO PARA ETIQUETAS BINARIAS

3.2. Retrieval Augmented Generation (RAG)

En la implementación del sistema recomendador, se emplea una base de datos vectorial que agiliza la recuperación de secciones de textos constitucionales altamente similares a la representación vectorial de la búsqueda del usuario.

3.2.1. Modificación a RAG. La técnica RAG recupera la información con la mayor similitud a la pregunta. En el contexto de textos constitucionales, RAG puede introducir al sistema recomendador textos con implicaciones jurídicas opuestas al documento buscado, ya que la similitud vectorial no garantiza una similitud jurídica. La adaptación de RAG se basa en la intervención de un agente experto entre la recuperación y la generación del modelo recomendador. Después de recuperar los K textos más similares a la *query* de la base de datos vectorial, el agente evalúa individualmente si estos textos son aceptables o no. Luego, entrega solo los textos cuya similitud jurídica con la *query* ha sido clasificada como INTACTA o MENOR. En resumen, el agente experto filtra la salida de RAG, proporcionando al modelo final solo los textos considerados aceptables en términos de similitud jurídica.

3.2.2. KDB.AI. A nivel técnico, se implementa una base de datos vectorial proporcionada por la empresa KDB.AI. A continuación, se presenta la estructura de la base de datos. En la metadata, se incluyen el tópico (*topic*) y la procedencia (*country*) del texto constitucional analizado (*text*). Además, la vectorización se presenta en *embeddings*, que consisten en vectores de 1536 dimensiones obtenidos a través del modelo GPT ada-002, con una búsqueda de similitud utilizando la distancia euclidiana (L2) y la estrategia de indexación simple *flat*:

```
rag_schema = {
  "columns": [
    {"name": "id",
      "pytype": "str"},
    {"name": "text",
      "pytype": "bytes"},
    {"name": "country",
      "pytype": "str"},
    {"name": "topic",
      "pytype": "str"},
    {
      "name": "embeddings",
      "pytype": "float32",
      "vectorIndex": {
        "dims": 1536,
```

```

        "metric": "L2",
        "type": "flat"}},
    ],
}

```

3.3. Constitutional Recommender

Utilizando los textos constitucionales obtenidos de la base de datos vectorial y sometidos a un proceso de filtrado por el agente experto previamente calibrado, se emplea un modelo basado en la tecnología GPT para formular una recomendación en respuesta a la cuestión planteada, teniendo en cuenta tanto los mencionados documentos como su metadatos.

La modelación clásica de un sistema recomendador se basa en la tríada user-item-interaction. En este contexto, un usuario sería una constitución, un ítem sería un tópico y la interacción sería el artículo de dicha constitución que trata sobre dicho tópico. A diferencia de un sistema clásico que se dedica a recomendar personalmente ítems dado interacciones, este sistema se basaría en recomendar usuarios dado interacciones, es decir, constituciones.

3.4. Dataset

El Proyecto Constitute es una iniciativa exhaustiva que recopila todas las constituciones que los países han adoptado a lo largo de su historia. En los experimentos realizados, se analizaron las constituciones de 53 países, como se ilustra en la Figura 4. Sólo se analizaron estas constituciones pues tienen la disponibilidad de una versión en español del texto en la base de datos. La mayoría de los países son de América (21), mientras que la segunda mayoría es Europa (16). También hay otros como Asia, África y Oceanía con 8, 7 y 1 constituciones respectivamente. En el caso Chileno, es importante señalar que, además de la constitución vigente, esta base de datos también cuenta con los proyectos de nueva constitución propuestos por la Convención Constitucional en 2022 y el Consejo Constitucional en 2023.

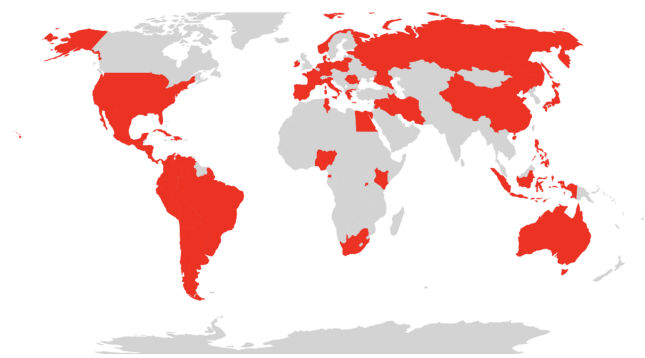


Figure 4. Países de las constituciones seleccionadas

Las oraciones de las constituciones están etiquetadas con tópicos provistos por el Proyecto Constitute. En total,

hay 325 tópicos que no necesariamente están presentes en todas las constituciones. Además, una constitución puede tener muchas oraciones asociadas a un mismo tópico. En promedio, un tópico se encuentra presente en casi 28 constituciones de las 53 analizadas.

En total, la información extraída del Proyecto Constitute corresponde a 54,624 oraciones que tienen como metadata un cierto tópico y una constitución de origen. Para posibilitar su rápida recuperación, estos datos fueron transformados a representaciones vectoriales mediante el modelo `ada-002` y almacenadas en una base de datos vectorial provista por `KDB.AI`.

4. Experimentos

Nuestra propuesta, basada en usar un agente para filtrar textos con similaridad jurídica, se va a compara con el *baseline* de utilizar únicamente la similaridad semántica de las representaciones vectoriales para recomendar textos.

4.1. Recomendación a la Propuesta del Consejo Constitucional 2023 en base a otras constituciones

El Instituto Res Publica identificó un subconjunto de 361 pares de oraciones que tienen una similaridad jurídica **IDÉNTICA** o **MENOR** entre la constitución vigente y la nueva propuesta. La tabla 4 muestra los resultados de las n recomendaciones a estos 361 textos analizados. Para la línea base, las recomendaciones son los n textos del tópico de la oración buscada con menos distancia euclidean. Para nuestro modelo, son los n textos con menor distancia euclidean y que además sean etiquetados por el agente como una relación jurídica aceptable. Para medir MAP y nDCG con el mismo criterio para ambos modelos, se procedió a catalogar con "1" a las recomendaciones de que coincidan con el texto de la constitución actual chilena que fue catalogada como **IDÉNTICA** o **MENOR** por el Instituto Res Publica.

n	Similaridad Jurídica (nuestro modelo)		Similaridad Semántica (baseline)	
	MAP	nDCG	MAP	nDCG
1	0.494	0.494	0.425	0.425
3	0.511	0.512	0.455	0.471
5	0.522	0.524	0.459	0.486
10	0.526	0.525	0.442	0.503

TABLE 4. MÉTRICAS MAP Y NDCG PARA LAS RECOMENDACIONES JURÍDICAS Y SEMÁNTICAS

Los resultados muestran que nuestro modelo tiende a tener mejores resultados que el *baseline* tanto con la métrica MAP como con nDCG. Este fenómeno puede atribuirse parcialmente al proceso de filtrado del agente, el cual excluye textos de las recomendaciones que, aunque puedan ser semánticamente similares, carecen de similitud jurídica. Este resultado conlleva a que otros textos con una mayor relevancia jurídica asciendan en prioridad, a pesar de presentar una similitud semántica inferior.

4.2. Países y Continentes Recomendados

Dados los resultados anteriores, se elige $n=5$ para visualizar las recomendaciones de otros países usando la similitud jurídica. Se eligió este tamaño dado que las métricas MAP y nDCG son marginalmente peores en comparación con $n=10$. Aparte del texto propuesto por el Consejo Constitucional, analiza también las recomendaciones al texto rechazado presentado por la Convención Constitucional en 2022.

Tal como muestra la Figura 5, la actual constitución del Consejo Constitucional es menos latinoamericanista que la propuesta de la Convención Constitucional de 2022. El texto anterior tiene un 61% de parecido con constituciones de otros países de América, mientras que la propuesta actual esta cifra baja a 47%. Además, la actual tiene mayor parecido a países europeos (19% contra 12%) y en menor medida a africanos (12% contra 8%).

Continente	N	Convención Constitucional 2022 (%)		Consejo Constitucional 2023 (%)	
		Suma	Promedio	Suma	Promedio
Chile	1	14.89	14.89	17.18	17.18
América	20	61.10	3.05	47.00	2.35
Europa	16	12.10	0.76	19.11	1.19
África	7	7.85	1.12	12.31	1.76
Asia	8	3.95	0.49	3.31	0.41
Oceanía	1	0.12	0.12	1.09	1.09
Total	53	100	-	100	-

TABLE 5. SIMILARIDAD JURÍDICA SEGÚN CONTINENTES

Dada la importancia de los países latinoamericanos, la Figura 5 visualiza el porcentaje de similitud jurídica de los textos resultantes de los últimos dos procesos constitucionales. Las principales diferencias entre ambos textos es que la rechazada propuesta de la Convención Constitucional de 2022 tenía un parecido significativamente mayor con las constituciones de Bolivia, Ecuador y Venezuela. En cambio, la propuesta del Consejo Constitucional en 2023 tiene un mayor parecido con la constición chilena vigente y tiende a tener menor parecido con las constituciones latinoamericanas. En específico, su similitud con Bolivia bajó un 7%, con Ecuador un 4% y con Venezuela un 3%.

4.3. Similitud de la Propuesta del Consejo Constitucional 2023 con otras constituciones chilenas

La tabla 6 muestra la similitud jurídica que tiene la propuesta de nueva constitución con tres textos constitucionales relevantes en la discusión política chilena: la fracasada propuesta de la Convención Constitucional en 2022, el borrador presentado por la presidenta Bachelet en 2017 y la actual Constitución Política de la República de Chile. A diferencia del experimento anterior, en este caso se comparó únicamente un par de constituciones, es decir, la propuesta con algún otro texto de Chile. Luego, para cada oración, se recuperó la semánticamente más parecida y se evaluó con el agente su similitud jurídica.

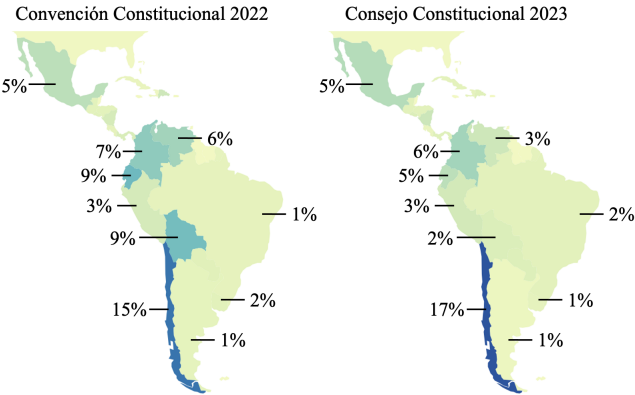


Figure 5. Similitud jurídica con países latinoamericanos

	Convención Constitucional 2022		Borrador Bachelet 2017		Constitución Actual 2005	
	N	%	N	%	N	%
INTACTO	12	1%	171	11%	226	15%
MENOR	162	11%	386	25%	419	27%
SUBTOTAL	174	11%	557	36%	645	42%
MAYOR	49	3%	67	4%	65	4%
SUPRESIÓN	1313	85%	912	59%	826	54%
SUBTOTAL	1362	89%	979	64%	891	58%
TOTAL	1536	100%	1536	100%	1536	100%

TABLE 6. SIMILARIDAD JURÍDICA DE LA PROPUESTA DEL CONSEJO CONSTITUCIONAL A OTROS TEXTOS CONSTITUCIONALES CHILENOS

Los resultados revelan que el texto que será sometido a plebiscito el 17 de diciembre de 2023 presenta 42% de similitud con la constitución actualmente vigente. Este hallazgo contrasta notablemente con la propuesta constitucional fallida de la Convención Constitucional de 2022, que muestra solo un 11% de similitud. En otro sentido, la constitución del Consejo Constitucional exhibe una similitud relativamente elevada con el borrador propuesto por la presidenta Bachelet en 2017, manteniendo un 36% de concordancia en cuanto a su contenido.

4.4. Sistema Recomendador de Contraste

Con el objetivo de comparar los resultados obtenidos por el recomendador actual, se decidió también entrenar un segundo modelo de recomendación basado en un método más tradicional basado en contenido. Este consiste en calcular similitud entre textos constitucionales a partir del calculo de la distancia coseno entre los vectores TF-idf de los documentos. Se asume similitud jurídica si las recomendaciones perteneces al mismo tópico del texto buscado.

Este modelo se entrenó sobre las representaciones vectoriales TF-idf de los artículos de textos constitucionales que contiene la base de datos de Constitue Project. El modelo, dada una constitución, recomienda por cada artículo de ella, los artículos de las constituciones con la que fue entrenada con la mayor similitud según distancia coseno.

Para evaluarla se usó como datos de prueba la actual constitución chilena, y se utilizaron las métricas MAP y nDCG para poder medir su rendimiento. Como se mencionó anteriormente, en este caso se definieron como recomendaciones relevantes aquellos artículos recomendados que comparten el mismo tópico que el texto al que se recomendó.

Los resultados obtenidos están dados por la tabla 7:

N	Similaridad	
	Vectores	TF-idf
	MAP	nDCG
1	0.261	0.261
3	0.315	0.403
5	0.306	0.460
10	0.287	0.551

TABLE 7. MÉTRICAS MAP Y NDCG PARA LA RECOMENDACIÓN BASADA EN CONTENIDO

Es interesante analizar que la métrica MAP empeora sus resultados cuando la lista de recomendaciones es mayor a $n=3$. En contraste, nDCG mejora sustantivamente sus resultados en la medida que aumenta el tamaño de la recomendación.

5. Conclusiones

El estudio y desarrollo de un sistema recomendador basado en la similaridad jurídica de textos constitucionales utilizando un modelo GPT *fine-tuned* como agente experto para filtrar recomendaciones ha demostrado ser una estrategia prometedora. Al contrastar este enfoque con un *baseline* que se apoya únicamente en la similaridad semántica, se observa un rendimiento superior en términos de métricas como MAP y nDCG. Este resultado destaca la importancia de considerar no solo la similitud en el significado general de las palabras, sino también la similitud jurídica, que puede ser crucial en el ámbito constitucional.

El análisis de la similaridad jurídica entre la propuesta de constitución del Consejo Constitucional en 2023 y otras constituciones chilenas revela cambios significativos en la orientación legal. La propuesta actual muestra una menor afinidad con las constituciones latinoamericanas, particularmente con Bolivia, Ecuador y Venezuela, en comparación con la propuesta rechazada de la Convención Constitucional en 2022. Además, se destaca un aumento en la similitud con la constitución chilena vigente y otros países europeos.

La implementación de un sistema recomendador que combina Retrieval-Augmented Generation (RAG) con un agente experto para filtrar recomendaciones abre nuevas perspectivas en la aplicación de modelos de lenguaje en el ámbito jurídico y constitucional. Este enfoque puede ser valioso no solo para el caso chileno, sino también como un marco general para evaluar la similitud jurídica en otros contextos legales. Este trabajo ofrece una contribución significativa al campo de los sistemas recomendadores aplicados a textos jurídicos y constitucionales, destacando la importancia de considerar la similitud jurídica para realizar recomendaciones más precisas y contextualmente relevantes en el ámbito legal.

6. Trabajo Futuro

Como trabajo futuro se propone utilizar otras métricas de distancia para el cálculo de similitud de representaciones vectoriales. En particular, la distancia coseno se presenta como una buena alternativa a la distancia euclídeana para capturar de mejor forma similaridades de vectores de alta dimensionalidad. Además, resulta interesante ampliar este trabajo a otros idiomas. En este estudio se analizaron únicamente constituciones escritas en español o que fueron traducidas por el equipo del Proyecto Constitute, por lo que sería valioso ver cómo se comportan las recomendaciones en el caso multilingüal. En relación a las representaciones vectoriales, se podrían probar otros embeddings, como por ejemplo, los entregados por modelos como BERT. De esta forma, se podrían ampliar los *baselines* analizados. Por último, los resultados prometedores en el área constitucional invitan a explorar otras áreas legales donde el uso de modelos GPT especializados puedan ser de aporte, como por ejemplo, casos judiciales o la redacción de leyes.

References

- [1] K. Bao, J. Zhang, Y. Zhang, W. Wang, F. Feng, and X. He, "Tallrec: An effective and efficient tuning framework to align large language model with recommendation," Sep. 2023. [Online]. Available: <http://dx.doi.org/10.1145/3604915.3608857>
- [2] Y. Gao, T. Sheng, Y. Xiang, Y. Xiong, H. Wang, and J. Zhang, "Chatrec: Towards interactive and explainable llms-augmented recommender system," 2023.
- [3] O. Shahmirzadi, A. Lugowski, and K. Younge, "Text similarity in vector space models: A comparative study," 2018.
- [4] N. Gahman and V. Elangovan, "A comparison of document similarity algorithms," 2023.
- [5] P. Bhattacharya, K. Ghosh, A. Pal, and S. Ghosh, "Methods for computing legal document similarity: A comparative study," 2020.
- [6] —, "Legal case document similarity: You need both network and text," 2022.
- [7] R. S. de Oliveira and E. G. S. Nascimento, "Analysing similarities between legal court documents using natural language processing approaches based on transformers," 2023.
- [8] S. Kumar, P. K. Reddy, V. B. Reddy, and A. Singh, "Similarity analysis of legal judgments," New York, NY, USA, 2011. [Online]. Available: <https://doi.org/10.1145/1980422.1980439>
- [9] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. tau Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," 2021.