

Tipo de artículo: Artículo originales

Temática: Inteligencia artificial

Recibido: dd/mm/aa | Aceptado: dd/mm/aa | Publicado: dd/mm/aa

Democratización de Herramientas de Inteligencia Artificial para la Gestión de Documentos PDF en Contextos con Recursos Limitados

Democratization of Artificial Intelligence Tools for PDF Document Management in Resource-Limited Contexts

Joaquin Rivas Sánchez **0009-0007-7971-8259**^{1*}

Naylin Brizuela Capote **0009-0000-5059-8248**¹

Angel Alberto Vazquez Sánchez **0000-0002-3130-7983**¹

¹Universidad de las Ciencias Informáticas. Km 2 ½ carretera a San Antonio de los Baños, Reparto Torrens, La Lisa, La Habana

*Autor para correspondencia: (joaquiners@estudiantes.uci.cu)

RESUMEN

Este artículo abordó la necesidad de desarrollar una herramienta accesible para la gestión y análisis de documentos PDF en entornos con recursos limitados, como el contexto cubano, donde las restricciones de conectividad y métodos de pago internacionales dificultan el acceso a soluciones avanzadas como ChatPDF y Humata.ai. Además, se consideró crucial avanzar en el desarrollo de tecnologías que promuevan la soberanía tecnológica, reduciendo la dependencia de plataformas extranjeras. Para solucionar este problema, se propuso un prototipo basado en la técnica Retrieval-Augmented Generation (RAG), empleando el modelo de lenguaje llama-3.2-3B y all-mpnet-base-v2 para la generación de word embeddings. La principal funcionalidad del prototipo incluye la capacidad de responder preguntas abiertas (open-domain question answering). Adicionalmente, el prototipo demostró su potencial para funcionar eficazmente en entornos con bajos recursos. Se concluye que este prototipo puede ser desarrollado aún más y tiene el potencial de democratizar el acceso a tecnologías avanzadas de inteligencia artificial, lo que fortalecería la capacidad local para gestionar documentos PDF de manera eficiente.

Palabras clave: inteligencia artificial; RAG; procesamiento de lenguaje natural; conectividad limitada; código abierto.

ABSTRACT

This article addressed the need to develop an accessible tool for managing and analyzing PDF documents in resource-limited environments, such as the Cuban context, where connectivity restrictions and international payment methods hinder access to advanced solutions like ChatPDF and Humata.ai. Furthermore, advancing the development of technologies that promote technological sovereignty and reduce reliance on foreign platforms was deemed crucial. To address this issue, a prototype was proposed based on the Retrieval-Augmented Generation (RAG) technique, utilizing the llama-3.2-3B language model and all-mpnet-base-v2 for generating word embeddings. The prototype's main functionality includes the ability to perform open-domain question answering. Additionally, the prototype demonstrated its potential to function effectively in low-resource settings. It is concluded that, with further development, this prototype has the potential to democratize access to advanced artificial intelligence technologies, strengthening local capabilities for efficient PDF document management.

Keywords: artificial intelligence; RAG; natural language processing; limited connectivity; open source.

Introducción

En las referencias solo se incluyen los trabajos citados explícitamente en el texto. Evite el uso de fuentes no confiables. Toda la bibliografía usada debe estar acotada en el cuerpo del documento siguiendo las pautas de las normas ISO 690 de la forma (Apellido, año). Ejemplos: de forma simple.

Los párrafos se escribirán en Times New Roman a 12 puntos, con espaciado 1.5, justificado y una línea en blanco como separador (usando solo enter como separador).

La introducción constituye una presentación del tema y debe incluir los objetivos trazados, exponer brevemente los trabajos más relevantes y destacar las contribuciones de otros autores al tema objeto de estudio, así como justificar las razones por las que se realiza la investigación.

Métodos o Metodología Computacional

Los párrafos se escribirán en Times New Roman a 12 puntos y con espaciado 1,5 y una línea en blanco como separador.

En esta sección se explica cómo se hizo la investigación. Se describe el diseño de la misma y se explica cómo se llevó a la práctica, justificando la elección de métodos y técnicas de forma tal que un lector pueda repetir el estudio.

Subepígrafes en caso de utilizarse

Los párrafos se escribirán en Times New Roman a 12 puntos y con espaciado 1,5 y una línea en blanco como separador.

SubSubepígrafes en caso de utilizarse

Los párrafos se escribirán en Times New Roman a 12 puntos y con espaciado 1,5 y una línea en blanco como separador.

Resultados y discusión

Los párrafos se escribirán en Times New Roman a 12 puntos y con espaciado 1,5 y una línea en blanco como separador.

Los resultados obtenidos se exponen después de explicar las técnicas seleccionadas y descritas en la sección anterior. Los mismos deben apoyarse por el uso de las tablas, figuras y diagramas, que expresan de forma clara los resultados del estudio realizado por los autores.

Se espera encontrar en esta sección aquellos elementos que hacen que la investigación presentada constituya una novedad o una mejora en su campo de acción y su superioridad con respecto a soluciones similares. Se debe enfatizar en la validación del resultado y su posible aplicación en contextos diferentes al reportado en el artículo. En la parte de la discusión se presenta el análisis de los resultados obtenidos que deben corresponder

a los objetivos planteados en el artículo y su contraste y comparación con lo reportado previamente en la literatura científica.

Cuando en el texto tenemos una relación en forma de lista, se utilizará forma numérica:

1. Texto.

a) Texto.

1) Texto.

b) Texto.

2. Texto.

Todas las tablas y figuras deberán ser mencionadas en el texto. A continuación vea una ejemplo de la Figura 1 y de la Tabla 1.

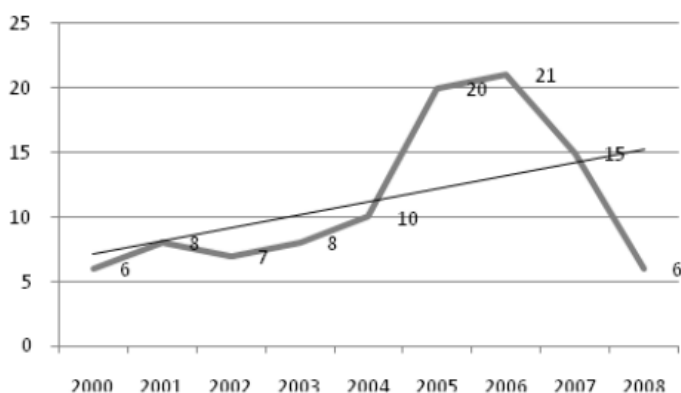


Fig. 1 - El título de las figuras se colocará en la parte inferior, centrado, utilizando numeración secuencial según el orden en que aparecen en el trabajo.

Si se desea usar alguna ecuación:

$$\text{Razón de tasa de mortalidad}_{\text{causa X}} = \frac{\text{Tasa de mortalidad masculina}_{\text{causa X}}}{\text{Tasa de mortalidad femenina}_{\text{causa X}}} \quad (1)$$

Tabla 1 - El título de las tablas en la parte inferior, centrado, utilizando numeración secuencial según el orden en que aparecen en el trabajo.

	Columna 1	Columna 2	Columna 3
Fila 1	x ^a	y	z
Fila 2	x	y	z ^b

^a De emplear notas aclaratorias se colocarán al pie de la tabla.

^b Otra nota aclaratoria.

Texto¹ Si se quiere hacer un llamado para una nota aclaratoria ² (este va en superíndice, pero NO entre paréntesis).

Conclusiones

Los párrafos se escribirán en Times New Roman a 12 puntos y con espaciado 1,5 y una línea en blanco como separador.

Las conclusiones se derivan del trabajo realizado. No son una repetición de incluido en el Resumen. Toda conclusión debe estar fundamentada en lo expuesto y discutido en el trabajo y debe reflejar el cumplimiento de los objetivos. Deben indicar cómo el trabajo contribuye o es un avance en el campo y objeto de estudio. Además, deben sugerir usos y trabajos futuros.

Agradecimientos (Opcional)

Los párrafos se escribirán en Times New Roman a 12 puntos y con espaciado 1,5 y una línea en blanco como separador.

Se añaden los nombres de personas que contribuyeron a la investigación pero que no se consideran como parte del colectivo de autores. Se incluyen los nombres de instituciones o proyectos que proporcionaron facilidades para la realización de la investigación tanto materiales, logísticas o financieras.

Referencias

En las referencias solo se incluyen los trabajos citados explícitamente en el texto. Evite el uso de fuentes no confiables. Toda la bibliografía usada debe estar (Toledo-Romaní et al., 2023) acotada en el cuerpo del documento siguiendo las pautas de las normas ISO 690 de la forma (Apellido, año). Ejemplos: forma simple.

Referencias

María Eugenia Toledo-Romaní, Mayra García-Carmenate, Carmen Valenzuela-Silva, Waldemar Baldoquín-Rodríguez, Marisel Martínez-Pérez, Meiby Rodríguez-González, Beatriz Paredes-Moreno, Ivis Mendoza-Hernández, Raúl González-Mujica Romero, Oscar Samón-Tabio, et al. Safety and efficacy of the two doses conjugated protein-based soberana-02 covid-19 vaccine and of a heterologous three-dose combination with soberana-plus: a double-blind, randomised, placebo-controlled phase 3 clinical trial. *The Lancet Regional Health–Americas*, 18, 2023.

Conflicto de interés

El autor autoriza la distribución y uso de su artículo. (12 Ptos. Just.)

Contribuciones de los autores

1. Conceptualización: Nombre y Apellidos del autor.
2. Curación de datos: Nombre y Apellidos del autor
3. Análisis formal: Nombre y Apellidos del autor
4. Adquisición de fondos: Nombre y Apellidos del autor
5. Investigación: Nombre y Apellidos del autor
6. Metodología: Nombre y Apellidos del autor

7. Administración del proyecto: Nombre y Apellidos del autor
8. Recursos: Nombre y Apellidos del autor
9. Software: Nombre y Apellidos del autor
10. Supervisión: Nombre y Apellidos del autor
11. Validación: Nombre y Apellidos del autor
12. Visualización: Nombre y Apellidos del autor
13. Redacción - borrador original: Nombre y Apellidos del autor
14. Redacción - revisión y edición: Nombre y Apellidos del autor

OBSERVACIÓN: Cada rol se define de la siguiente forma:

- Conceptualización - Ideas; formulación o evolución de los objetivos y metas generales de la investigación.
- Curación de datos - Actividades de gestión para anotar (producir metadatos), depurar datos y mantener los datos de la investigación (incluido el código de software, cuando sea necesario para interpretar los propios datos) para su uso inicial y su posterior reutilización.
- Análisis formal - Aplicación de técnicas estadísticas, matemáticas, computacionales u otras técnicas formales para analizar o sintetizar datos de estudio.
- Adquisición de fondos - Adquisición del apoyo financiero para el proyecto que conduce a esta publicación.
- Investigación - Realización de una investigación y proceso de investigación, realizando específicamente los experimentos, o la recolección de datos/evidencia.
- Metodología - Desarrollo o diseño de la metodología; creación de modelos.
- Administración del proyecto - Responsabilidad de gestión y coordinación de la planificación y ejecución de la actividad de investigación.

- Recursos - Suministro de materiales de estudio, reactivos, materiales, pacientes, muestras de laboratorio, animales, instrumentación, recursos informáticos u otras herramientas de análisis.
- Software - Programación, desarrollo de software; diseño de programas informáticos; implementación del código informático y de los algoritmos de apoyo; prueba de los componentes de código existentes.
- Supervisión - Responsabilidad de supervisión y liderazgo en la planificación y ejecución de actividades de investigación, incluyendo la tutoría externa al equipo central.
- Validación - Verificación, ya sea como parte de la actividad o por separado, de la replicabilidad/reproducción general de los resultados/experimentos y otros productos de la investigación.
- Visualización - Preparación, creación y/o presentación del trabajo publicado, específicamente la visualización/presentación de datos.
- Redacción - borrador original - Preparación, creación y/o presentación del trabajo publicado, específicamente la redacción del borrador inicial (incluyendo la traducción sustantiva).
- Redacción - revisión y edición - Preparación, creación y/o presentación del trabajo publicado por los miembros del grupo de investigación original, específicamente revisión crítica, comentario o revisión - incluyendo las etapas previas o posteriores a la publicación.

Financiación

Quién financia la investigación. (12 Ptos. Just.)

Notes

¹Notas aclaratorias del texto, estas estarán ubicadas al final del trabajo.

²Esta es otra nota al final

Observaciones: - No puede haber quiebras en los títulos. - Las tablas complejas o con celdas combinadas deben ser trabajadas como imágenes.