

Anexo Emergencias y Ayuda Sanitaria.

Para este experimento se utilizó el manual Guías INSARAG Volumen II - Manual B

url del manual: <https://www.insarag.org/methodology/insarag-guidelines/>

Objetivo del Documento

El documento "Guías INSARAG Volumen II - Manual B" tiene como objetivo principal proporcionar orientación sobre la coordinación operacional y la preparación de los equipos de búsqueda y rescate urbano (USAR) para misiones tanto nacionales como internacionales. Este manual está dirigido a la coordinación operacional de INSARAG, al equipo administrativo de USAR y a la Secretaría de INSARAG, estableciendo estándares mínimos y capacidades requeridas para las operaciones de coordinación .

Resumen del Contenido

El resumen del documento describe que se enfoca en la capacitación, preparación y coordinación de equipos USAR. Se basa en estándares internacionales y detalla el ciclo de respuesta internacional de USAR, además de las responsabilidades clave en operaciones de búsqueda y rescate, como la coordinación con Naciones Unidas y los países afectados.

El documento "Guías INSARAG Volumen II - Manual B" tiene las siguientes estadísticas:

- **Número total de páginas:** 60
- **Número total de palabras:** 50,599
- **Número total de secciones (basado en líneas):** 7,062

Explicación de las Métricas. Se realiza una explicación de las métricas utilizadas para analizar cada respuesta obtenida mediante la generación en cada implementación.

Implementaciones. Se especifican las implementaciones y se indica el archivo con su respectivo código fuente.

Experimento.

Se realiza una aclaración respecto a la estructura de las preguntas y respuestas.

Se detallan de manera enumerada las preguntas realizadas en la plataforma de LE CHAT MISTRAL y sus correspondientes respuestas.

Luego se detalla la respuesta esperada y la respuesta generada, así como las métricas obtenidas y el análisis de las mismas para cada implementación.

Explicación de las Métricas

ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

Las métricas ROUGE miden la superposición de n-gramas entre la respuesta generada y la respuesta de referencia (esperada). Se utilizan comúnmente para evaluar la calidad de resúmenes y textos generados.

1. ROUGE-1:

- **Precision:** Mide la proporción de unigramas (palabras individuales) en la respuesta generada que también están en la respuesta de referencia.
- **Recall:** Mide la proporción de unigramas en la respuesta de referencia que también están en la respuesta generada.
- **F1-score:** Es la media armónica entre la precisión y el recall, proporcionando una medida balanceada de ambos.

2. ROUGE-2:

- **Precision:** Mide la proporción de bigramas (pares de palabras consecutivas) en la respuesta generada que también están en la respuesta de referencia.
- **Recall:** Mide la proporción de bigramas en la respuesta de referencia que también están en la respuesta generada.
- **F1-score:** Es la media armónica entre la precisión y el recall para bigramas.

3. ROUGE-L:

- **Precision:** Mide la proporción de la subsecuencia más larga común (LCS) en la respuesta generada que también está en la respuesta de referencia.
- **Recall:** Mide la proporción de la LCS en la respuesta de referencia que también está en la respuesta generada.
- **F1-score:** Es la media armónica entre la precisión y el recall para la LCS.

BERTScore

BERTScore es una métrica basada en embeddings que utiliza modelos de lenguaje preentrenados como BERT para evaluar la similitud semántica entre la respuesta generada y la respuesta de referencia.

- **Precision:** Mide la proporción de tokens en la respuesta generada que tienen una alta similitud con los tokens en la respuesta de referencia.
- **Recall:** Mide la proporción de tokens en la respuesta de referencia que tienen una alta similitud con los tokens en la respuesta generada.
- **F1:** Es la media armónica entre la precisión y el recall, proporcionando una medida balanceada de ambos.

Implementaciones.

Todas las implementaciones se realizaron en Python sobre Google Colab.

Preguntas y respuestas generadas con:

- **Implementación 1: RAG Básico.**

Todos los detalles de esta implementación pueden analizarse en el archivo

RagBasicoConPymuPDFEmergencias.ipynb

- **Implementación 2: RAG con Haystack**

Todos los detalles de esta implementación pueden analizarse en el archivo

RagConHaystackconPymuPDFEmergencias.ipynb

- **Implementación 3: RAG con LlamaIndex**

Todos los detalles de esta implementación pueden analizarse en el archivo

RAGconLLamaIndexconPymuPDFEmergencias.ipynb

- **EmergenciasMetricasLeChat.ipynb**

Se implementa el cálculo de las métricas de las respuestas generadas por LECHAT, es decir, son las métricas SIN RAG.

Experimento.

Observación: estructura de las preguntas y respuestas con la plataforma de LE CHAT MISTRAL.

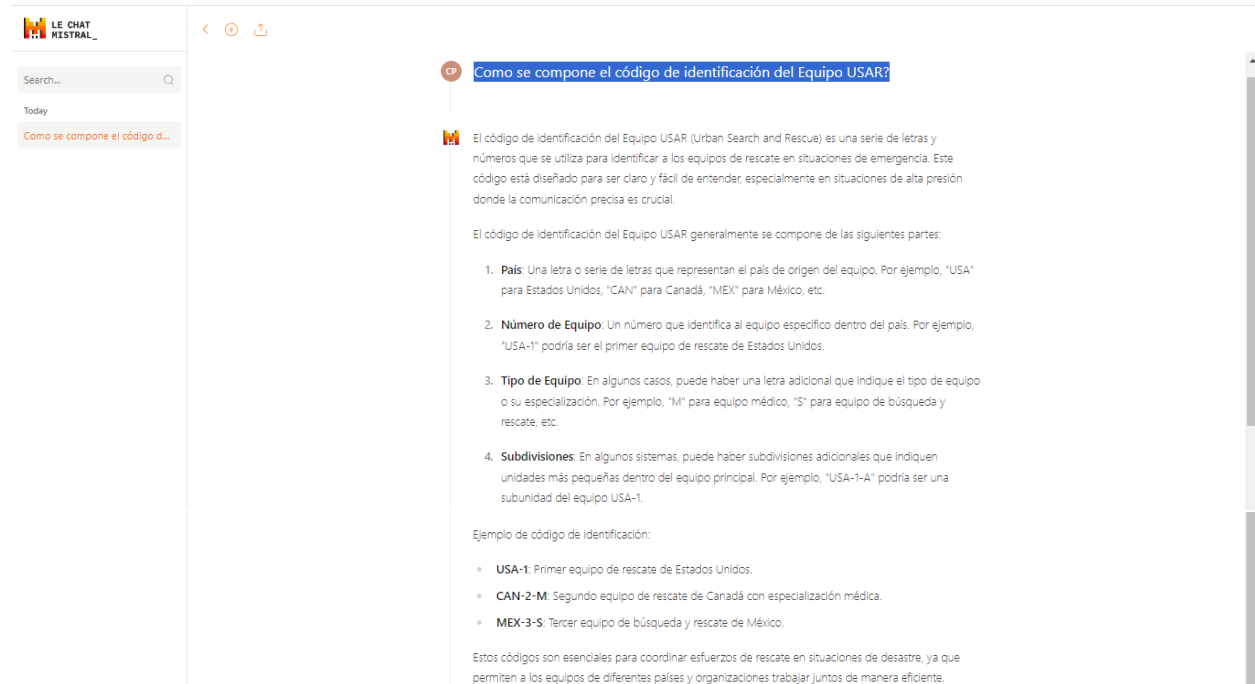
Se realizaron las mismas preguntas a la plataforma LE CHAT con las que se evaluaron con el sistema RAG en sus tres implementaciones. En cada respuesta se puede apreciar la generalización en la respuesta debido a que el LLM no posee información exacta del contexto en cuestión (Guías INSARAG Volumen II - Manual B) pero si posee información general acerca de esta guía.

En este experimento se detallan las preguntas y capturas de pantalla de las respuestas obtenidas por la plataforma LECHAT. Luego las respuestas generadas en cada implementación. Aclaración: la respuesta esperada en cada implementación es la misma para poder realizar una comparación exacta de esta con cada respuesta generada obteniendo de estas las respectivas métricas. Luego por cada respuesta se detallan los valores obtenidos de las métricas utilizadas para evaluar las respuestas obtenidas.

Preguntas y respuestas del experimento.

Pregunta 1: ¿Cómo se compone el código de identificación del Equipo USAR?

Respuesta generada por LECHAT



The screenshot shows a chat window titled "LE CHAT MISTRAL...". On the left is a search bar and a list of messages, with the selected message being "Como se compone el código d...". The main chat area displays a question in a blue header: "Como se compone el código de identificación del Equipo USAR?". The answer, generated by "MISTRAL", explains that the USAR code is a series of letters and numbers used for emergency rescue teams. It lists four components: 1. País (Country), 2. Número de Equipo (Team Number), 3. Tipo de Equipo (Team Type), and 4. Subdivisiones (Subdivisions). Examples provided are USA-1, CAN-2-M, and MEX-3-S. A note at the bottom states that these codes are essential for coordinating rescue efforts in disaster situations.

Como se compone el código de identificación del Equipo USAR?

El código de identificación del Equipo USAR (Urban Search and Rescue) es una serie de letras y números que se utiliza para identificar a los equipos de rescate en situaciones de emergencia. Este código está diseñado para ser claro y fácil de entender, especialmente en situaciones de alta presión donde la comunicación precisa es crucial.

El código de identificación del Equipo USAR generalmente se compone de las siguientes partes:

1. **País:** Una letra o serie de letras que representan el país de origen del equipo. Por ejemplo, "USA" para Estados Unidos, "CAN" para Canadá, "MEX" para México, etc.
2. **Número de Equipo:** Un número que identifica al equipo específico dentro del país. Por ejemplo, "USA-1" podría ser el primer equipo de rescate de Estados Unidos.
3. **Tipo de Equipo:** En algunos casos, puede haber una letra adicional que indique el tipo de equipo o su especialización. Por ejemplo, "M" para equipo médico, "S" para equipo de búsqueda y rescate, etc.
4. **Subdivisiones:** En algunos sistemas, puede haber subdivisiones adicionales que indiquen unidades más pequeñas dentro del equipo principal. Por ejemplo, "USA-1-A" podría ser una subunidad del equipo USA-1.

Ejemplo de código de identificación:

- **USA-1:** Primer equipo de rescate de Estados Unidos.
- **CAN-2-M:** Segundo equipo de rescate de Canadá con especialización médica.
- **MEX-3-S:** Tercer equipo de búsqueda y rescate de México.

Estos códigos son esenciales para coordinar esfuerzos de rescate en situaciones de desastre, ya que permiten a los equipos de diferentes países y organizaciones trabajar juntos de manera eficiente.

Transcripción.

El código de identificación del Equipo USAR (Urban Search and Rescue) es una serie de letras y números que se utiliza para identificar a los equipos de rescate en situaciones de emergencia. Este código está diseñado para ser claro y fácil de entender, especialmente en situaciones de alta presión donde la comunicación precisa es crucial.

El código de identificación del Equipo USAR generalmente se compone de las siguientes partes:

1. País: Una letra o serie de letras que representan el país de origen del equipo. Por ejemplo, "USA" para Estados Unidos, "CAN" para Canadá, "MEX" para México, etc.
2. Número de Equipo: Un número que identifica al equipo específico dentro del país. Por ejemplo, "USA-1" podría ser el primer equipo de rescate de Estados Unidos.
3. Tipo de Equipo: En algunos casos, puede haber una letra adicional que indique el tipo de equipo o su especialización. Por ejemplo, "M" para equipo médico, "S" para equipo de búsqueda y rescate, etc.
4. Subdivisiones: En algunos sistemas, puede haber subdivisiones adicionales que indiquen unidades más pequeñas dentro del equipo principal. Por ejemplo, "USA-1-A" podría ser una subunidad del equipo USA-1.

Ejemplo de código de identificación:

- USA-1: Primer equipo de rescate de Estados Unidos.
- CAN-2-M: Segundo equipo de rescate de Canadá con especialización médica.
- MEX-3-S: Tercer equipo de búsqueda y rescate de México.

Estos códigos son esenciales para coordinar esfuerzos de rescate en situaciones de desastre, ya que permiten a los equipos de diferentes países y organizaciones trabajar juntos de manera eficiente.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.19310344827586207,
recall=0.5894736842105263, fmeasure=0.2909090909090909)
ROUGE-2 score: Score(precision=0.0726643598615917,
recall=0.22340425531914893, fmeasure=0.10966057441253263)
ROUGE-L score: Score(precision=0.12413793103448276,
recall=0.37894736842105264, fmeasure=0.18701298701298702)
BERTScore - Precision: 0.6355971693992615, Recall: 0.6993028521537781, F1:
0.6659298539161682
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.193 indica que el 19.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.589 sugiere que el sistema recupera el 58.9% de las palabras relevantes, lo que refleja un **recall moderado** pero una **precisión relativamente baja**.
- **F1-Score:** 0.291 refleja un **equilibrio modesto** entre precisión y recall, lo que sugiere que el sistema captura una cantidad razonable de contenido relevante pero también genera contenido irrelevante.

2. ROUGE-2 Score:

- **Precision:** 0.073 indica que solo el 7.3% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.223 sugiere que el sistema recupera el 22.3% de los bigrams relevantes.
- **F1-Score:** 0.110 refleja un **rendimiento bajo** en la captura de secuencias de dos palabras, lo que indica dificultades en generar relaciones coherentes entre palabras clave.

3. ROUGE-L Score:

- **Precision:** 0.124 indica que el 12.4% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.379 muestra que el sistema recupera el 37.9% de la secuencia relevante más larga.
- **F1-Score:** 0.187 refleja un **rendimiento moderado-bajo** en la captura de secuencias estructuradas, con una **precisión limitada** y un **recall moderado**.

4. BERTScore:

- **Precision:** 0.636 indica que el 63.6% de las palabras generadas son semánticamente relevantes.

- **Recall:** 0.699 sugiere que el sistema captura el 69.9% del contenido semántico relevante.
- **F1-Score:** 0.666 refleja un **rendimiento semántico razonable**, con un **buen balance** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **recall moderado** pero una **precisión baja**, lo que sugiere que genera una cantidad considerable de contenido irrelevante. El rendimiento en secuencias (ROUGE-2 y ROUGE-L) es bajo, lo que indica dificultades en la captura de relaciones coherentes entre palabras clave y en generar respuestas estructuradas. El **BERTScore** muestra un **rendimiento semántico razonable**, lo que sugiere que el sistema capta bien el significado general, pero aún necesita mejorar en **precisión literal** y en la **generación de secuencias coherentes** para proporcionar respuestas más relevantes y exactas.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

La Identificación del Equipo está compuesta de dos partes: las tres letras del Código Olímpico del país de origen del equipo y un número de dos dígitos para diferenciar equipos del mismo país. Los equipos multinacionales como ONGs no usan el Código Olímpico del país de origen del equipo sino las letras "SAR" para ser identificados. Los designadores de equipos 01 a 09 son para ser usados con equipos clasificados IEC/IER, los designadores 10 a 99 son para ser usados con equipos no clasificados.

Respuesta generada con RAG Básico:

El código de identificación del Equipo USAR se compone de dos partes:

1. ****Tres letras del Código Olímpico del país de origen del equipo**:** Esta parte identifica el país de procedencia del equipo.
2. ****Un número de dos dígitos**:** Este número se utiliza para diferenciar equipos del mismo país.

Para equipos multinacionales como ONGs, en lugar del Código Olímpico del país de origen, se utilizan las letras "SAR" para su identificación.

Además, los designadores de equipos del 01 al 09 se utilizan para equipos clasificados IEC/IER, mientras que los designadores del 10 al 99 se utilizan para equipos no clasificados. La Coordinación Operacional INSARAG administra estos designadores de equipo de país y actualiza el directorio USAR en www.insarag.org.

Métricas:

ROUGE-1 score: Score(precision=0.5579710144927537,
recall=0.8105263157894737, fmeasure=0.6609442060085836)
ROUGE-2 score: Score(precision=0.40875912408759124,
recall=0.5957446808510638, fmeasure=0.48484848484848486)
ROUGE-L score: Score(precision=0.5144927536231884,
recall=0.7473684210526316, fmeasure=0.6094420600858369)
BERTScore - Precision: 0.7877885699272156, Recall: 0.8694505095481873, F1:
0.8266075849533081

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.558 indica que el 55.8% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.811 sugiere que el sistema recupera el 81.1% de las palabras relevantes, lo que refleja un **recall alto**.
- **F1-Score:** 0.661 muestra un **buen equilibrio** entre precisión y recall, lo que indica que el sistema genera contenido relevante de manera eficiente.

2. ROUGE-2 Score:

- **Precision:** 0.409 indica que el 40.9% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.596 muestra que el sistema recupera el 59.6% de los bigrams relevantes.
- **F1-Score:** 0.485 refleja un **buen rendimiento** en la generación de secuencias de palabras, con una **buena capacidad de recuperación** de relaciones entre palabras clave.

3. ROUGE-L Score:

- **Precision:** 0.514 indica que el 51.4% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.747 muestra que el sistema recupera el 74.7% de la secuencia relevante más larga.
- **F1-Score:** 0.609 refleja un **buen rendimiento** en la captura de secuencias estructuradas, con un **alto recall** y una **precisión razonable**.

4. BERTScore:

- **Precision:** 0.788 indica que el 78.8% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.869 sugiere que el sistema captura el 86.9% del contenido semántico relevante.
- **F1-Score:** 0.827 refleja un **muy buen equilibrio** entre precisión y recall en términos de correspondencia semántica.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican que el sistema tiene un **buen rendimiento** tanto en la generación de contenido relevante como en la captura de secuencias de palabras significativas. El **recall alto** y la **precisión razonable** muestran que el sistema es capaz de recuperar gran parte del contenido relevante sin generar demasiado contenido irrelevante. El **BERTScore** refuerza la idea de que el sistema es **fuerte en términos de semántica**, capturando

el significado general de manera efectiva. En general, el sistema es **sólido tanto en precisión como en recall**, con un buen rendimiento en la generación de respuestas coherentes y significativas.

Respuesta generada con RAG con Haystack:

El código de identificación del Equipo USAR se compone de dos partes: las tres letras del Código Olímpico del país de origen del equipo y un número de dos dígitos para diferenciar equipos del mismo país. Los equipos multinacionales, como ONGs, no usan el Código Olímpico del país de origen del equipo, sino las letras "SAR" para ser identificados. Los designadores de equipos 01 a 09 son para equipos clasificados IEC/IER, mientras que los designadores 10 a 99 son para equipos no clasificados. La Coordinación Operacional INSARAG administra los designadores de equipo de país y actualiza el directorio USAR en www.insarag.org.

Métricas:

```
ROUGE-1 score: Score(precision=0.7435897435897436,  
recall=0.9157894736842105, fmeasure=0.820754716981132)  
ROUGE-2 score: Score(precision=0.6982758620689655,  
recall=0.8617021276595744, fmeasure=0.7714285714285714)  
ROUGE-L score: Score(precision=0.7350427350427351,  
recall=0.9052631578947369, fmeasure=0.8113207547169812)  
BERTScore - Precision: 0.858389139175415, Recall: 0.924645185470581, F1:  
0.8902861475944519
```

Análisis de las Métricas

- **ROUGE-1 Score:**

- **Precision:** 0.744 indica que el 74.4% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.916 sugiere que el sistema recupera el 91.6% de las palabras relevantes, lo que refleja un **recall muy alto**.
- **F1-Score:** 0.821 muestra un **excelente equilibrio** entre precisión y recall, lo que indica que el sistema es altamente eficiente en la generación de contenido relevante.

- **ROUGE-2 Score:**

- **Precision:** 0.698 indica que el 69.8% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.862 muestra que el sistema recupera el 86.2% de los bigrams relevantes.
- **F1-Score:** 0.771 refleja un **rendimiento sólido** en la captura de secuencias significativas de palabras, con una **alta capacidad de recuperación y precisión**.

- **ROUGE-L Score:**

- **Precision:** 0.735 indica que el 73.5% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.905 muestra que el sistema recupera el 90.5% de la secuencia relevante más larga.
 - **F1-Score:** 0.811 refleja un **alto rendimiento** en la captura de secuencias estructuradas, con una **precisión sólida** y **alto recall**.
- **BERTScore:**
 - **Precision:** 0.858 indica que el 85.8% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.925 sugiere que el sistema captura el 92.5% del contenido semántico relevante.
 - **F1-Score:** 0.890 refleja un **rendimiento excepcional** en términos de correspondencia semántica, con un **muy buen equilibrio** entre precisión y recall.

Respuesta generada con RAG con LLaIndex

El código de identificación del Equipo USAR se compone de dos partes: las tres letras del Código Olímpico del país de origen del equipo y un número de dos dígitos para diferenciar equipos del mismo país. Los equipos multinacionales, como las ONGs, utilizan las letras "SAR" en lugar del Código Olímpico. Los designadores de equipos del 01 al 09 se usan para equipos clasificados IEC/IER, mientras que los designadores del 10 al 99 se usan para equipos no clasificados.

Métricas

```
ROUGE-1 score: Score(precision=0.8111111111111111,
recall=0.7684210526315789, fmeasure=0.7891891891891891)
ROUGE-2 score: Score(precision=0.6067415730337079,
recall=0.574468085106383, fmeasure=0.5901639344262296)
ROUGE-L score: Score(precision=0.7222222222222222,
recall=0.6842105263157895, fmeasure=0.7027027027027027)
BERTScore - Precision: 0.9123429656028748, Recall: 0.9138931632041931, F1:
0.9131174087524414
```

Análisis de las Métricas

1. **ROUGE-1 Score:**
 - **Precision:** 0.811 indica que el 81.1% de las palabras generadas son relevantes en comparación con el texto de referencia.
 - **Recall:** 0.768 muestra que el sistema recupera el 76.8% de las palabras relevantes, lo que refleja un **buen equilibrio** entre precisión y recall.
 - **F1-Score:** 0.789 indica un **rendimiento sólido** en la generación de contenido relevante, con una **alta precisión** y un **buen recall**.
2. **ROUGE-2 Score:**

- **Precision:** 0.607 indica que el 60.7% de los bigrams generados coinciden con el texto de referencia.
 - **Recall:** 0.574 muestra que el sistema recupera el 57.4% de los bigrams relevantes.
 - **F1-Score:** 0.590 refleja un **buen rendimiento** en la captura de secuencias significativas, aunque con espacio para mejorar el recall de las secuencias.
3. **ROUGE-L Score:**
- **Precision:** 0.722 indica que el 72.2% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.684 sugiere que el sistema recupera el 68.4% de la secuencia relevante más larga.
 - **F1-Score:** 0.703 refleja un **buen equilibrio** entre precisión y recall en términos de secuencias estructuradas.
4. **BERTScore:**
- **Precision:** 0.912 indica que el 91.2% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.914 muestra que el sistema captura el 91.4% del contenido semántico relevante.
 - **F1-Score:** 0.913 refleja un **rendimiento excepcional** en términos de correspondencia semántica, con un **equilibrio perfecto** entre precisión y recall.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **rendimiento sólido**, con un **buen equilibrio** entre precisión y recall tanto en la generación de palabras como en la captura de secuencias significativas. El **BERTScore** sugiere que el sistema es **excepcional en términos de semántica**, capturando de manera precisa y exhaustiva el significado del texto. En general, el sistema es muy **efectivo en la generación de contenido relevante**, tanto en términos de literalidad como de significado.

Pregunta 2: ¿En qué niveles se dividen las operaciones USAR?

Respuesta generada por LECHAT

The screenshot shows a chat window with a search bar at the top containing the text "En que niveles se dividen las operaciones USAR?". Below the search bar, there is a message from the user: "Como se compone el código d...". The chat interface shows a response from the system, LECHAT, which is a detailed explanation of the levels of Urban Search and Rescue (USAR) operations. The response is structured as follows:

Las operaciones de Urban Search and Rescue (USAR) se dividen en varios niveles para asegurar una respuesta eficiente y efectiva en situaciones de emergencia. Estos niveles están diseñados para proporcionar una estructura clara y organizada que permita a los equipos de rescate operar de manera coordinada. A continuación se describen los niveles típicos de operaciones USAR:

1. **Nivel Local (Local Level):**
 - **Equipos de Respuesta Inicial:** Estos equipos están compuestos por personal local, como bomberos, policías y paramédicos, que son los primeros en responder a una emergencia.
 - **Recursos Locales:** Incluyen equipos y suministros disponibles en la comunidad local, como vehículos de emergencia, equipos de rescate y hospitales locales.
2. **Nivel Estatal/Regional (State/Regional Level):**
 - **Equipos de Respuesta Estatal:** Estos equipos están formados por personal de agencias estatales y pueden incluir equipos de rescate especializados, como equipos de búsqueda y rescate, equipos de respuesta a desastres y equipos médicos.
 - **Recursos Estatales:** Incluyen equipos y suministros adicionales que pueden ser movilizadas desde otras partes del estado o región, como helicópteros, equipos de comunicación y

<div><div>Search...</div><div>Today</div><div>Como se compone el código d...</div></div>	<p>suministros médicos.</p> <p>3. Nivel Nacional (National Level):</p> <ul style="list-style-type: none">• Equipos de Respuesta Nacional: Estos equipos están formados por personal de agencias nacionales, como la Agencia Federal para el Manejo de Emergencias (FEMA) en Estados Unidos. Incluyen equipos de rescate altamente especializados y entrenados.• Recursos Nacionales: Incluyen equipos y suministros que pueden ser movilizadas desde cualquier parte del país, como equipos de rescate técnico, equipos de respuesta a desastres naturales y equipos médicos avanzados. <p>4. Nivel Internacional (International Level):</p> <ul style="list-style-type: none">• Equipos de Respuesta Internacional: Estos equipos están formados por personal de organizaciones internacionales, como las Naciones Unidas, la Cruz Roja Internacional y equipos de rescate de otros países.• Recursos Internacionales: Incluyen equipos y suministros que pueden ser movilizadas desde otros países, como equipos de rescate especializados, equipos médicos y suministros de ayuda humanitaria. <p>Cada nivel tiene sus propias capacidades y responsabilidades, y la coordinación entre estos niveles es crucial para una respuesta efectiva. En situaciones de emergencia, los equipos de respuesta inicial locales son generalmente los primeros en llegar a la escena, seguidos por equipos de nivel estatal, nacional e internacional según sea necesario. La estructura de niveles asegura que los recursos y el personal adecuados estén disponibles para manejar cualquier tipo de emergencia, desde incidentes locales menores hasta desastres a gran escala.</p>
--	--

Transcripción.

Las operaciones de Urban Search and Rescue (USAR) se dividen en varios niveles para asegurar una respuesta eficiente y efectiva en situaciones de emergencia. Estos niveles están diseñados para proporcionar una estructura clara y organizada que permita a los equipos de rescate operar de manera coordinada. A continuación se describen los niveles típicos de operaciones USAR:

1. Nivel Local (Local Level):

- **Equipos de Respuesta Inicial:** Estos equipos están compuestos por personal local, como bomberos, policías y paramédicos, que son los primeros en responder a una emergencia.
- **Recursos Locales:** Incluyen equipos y suministros disponibles en la comunidad local, como vehículos de emergencia, equipos de rescate y hospitales locales.

2. Nivel Estatal/Regional (State/Regional Level):

- **Equipos de Respuesta Estatal:** Estos equipos están formados por personal de agencias estatales y pueden incluir equipos de rescate especializados, como equipos de búsqueda y rescate, equipos de respuesta a desastres y equipos médicos.
- **Recursos Estatales:** Incluyen equipos y suministros adicionales que pueden ser movilizadas desde otras partes del estado o región, como helicópteros, equipos de comunicación y suministros médicos.

3. Nivel Nacional (National Level):

- **Equipos de Respuesta Nacional:** Estos equipos están formados por personal de agencias nacionales, como la Agencia Federal para el Manejo de Emergencias (FEMA) en Estados Unidos. Incluyen equipos de rescate altamente especializados y entrenados.
- **Recursos Nacionales:** Incluyen equipos y suministros que pueden ser movilizadas desde cualquier parte del país, como equipos de rescate técnico, equipos de respuesta a desastres naturales y equipos médicos avanzados.

4. Nivel Internacional (International Level):

- **Equipos de Respuesta Internacional:** Estos equipos están formados por personal de organizaciones internacionales, como las Naciones Unidas, la Cruz Roja Internacional y equipos de rescate de otros países.
- **Recursos Internacionales:** Incluyen equipos y suministros que pueden ser movilizadas desde otros países, como equipos de rescate especializados, equipos médicos y suministros de ayuda humanitaria.

Cada nivel tiene sus propias capacidades y responsabilidades, y la coordinación entre estos niveles es crucial para una respuesta efectiva. En situaciones de emergencia, los equipos de respuesta inicial locales son generalmente los primeros en llegar a la escena, seguidos por equipos de nivel estatal, nacional e internacional según sea necesario. La estructura de niveles asegura que los recursos y el personal adecuados estén disponibles para manejar cualquier tipo de emergencia, desde incidentes locales menores hasta desastres a gran escala.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.085995085995086, recall=0.603448275862069, fmeasure=0.15053763440860216)
ROUGE-2 score: Score(precision=0.027093596059113302, recall=0.19298245614035087, fmeasure=0.047516198704103674)
ROUGE-L score: Score(precision=0.0687960687960688, recall=0.4827586206896552, fmeasure=0.12043010752688173)
BERTScore - Precision: 0.6182373762130737, Recall: 0.7167237997055054, F1: 0.6638476848602295
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.086 indica que solo el 8.6% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.603 muestra que el sistema recupera el 60.3% de las palabras relevantes, lo que sugiere un **recall alto** pero con una **precisión muy baja**.
- **F1-Score:** 0.151 refleja un **desequilibrio significativo** entre precisión y recall, lo que indica que el sistema genera mucho contenido irrelevante.

2. ROUGE-2 Score:

- **Precision:** 0.027 indica que solo el 2.7% de los bigrams generados coinciden con el texto de referencia.

- **Recall:** 0.193 sugiere que el sistema recupera el 19.3% de los bigrams relevantes.
 - **F1-Score:** 0.048 refleja un **rendimiento muy bajo** en la captura de secuencias de palabras significativas, con una **precisión extremadamente baja**.
3. **ROUGE-L Score:**
- **Precision:** 0.069 indica que el 6.9% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.483 sugiere que el sistema recupera el 48.3% de la secuencia relevante más larga.
 - **F1-Score:** 0.120 refleja un **rendimiento bajo** en la captura de secuencias estructuradas, con una **baja precisión** pero un **recall moderado**.
4. **BERTScore:**
- **Precision:** 0.618 indica que el 61.8% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.717 muestra que el sistema captura el 71.7% del contenido semántico relevante.
 - **F1-Score:** 0.664 refleja un **rendimiento semántico moderado**, con un equilibrio aceptable entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** sugieren que el sistema tiene un **recall moderado a alto** pero una **precisión extremadamente baja**, lo que significa que genera una cantidad considerable de contenido irrelevante. El rendimiento en secuencias (ROUGE-2 y ROUGE-L) es especialmente bajo, lo que indica dificultades en generar relaciones significativas entre palabras y secuencias coherentes. El **BERTScore** muestra un **rendimiento semántico moderado**, lo que sugiere que el sistema capta razonablemente bien el significado general, pero necesita mejorar significativamente en **precisión literal** y en la **estructura de las secuencias** para generar respuestas más concisas y relevantes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

Las operaciones USAR se dividen en los cinco niveles listados en la tabla a continuación:

Nivel 1: Evaluación general del área.

Nivel 2: Evaluación del Triage del sitio de trabajo.

Nivel 3: Búsqueda y rescate rápido.

Nivel 4: Búsqueda y rescate total.

Nivel 5: Búsqueda y recuperación con cobertura total.

Respuesta generada con RAG Básico:

Las operaciones USAR se dividen en cinco niveles.

Métricas:

ROUGE-1 score: Score(precision=1.0, recall=0.13793103448275862, fmeasure=0.2424242424242424)

ROUGE-2 score: Score(precision=0.8571428571428571, recall=0.10526315789473684, fmeasure=0.1875)

ROUGE-L score: Score(precision=1.0, recall=0.13793103448275862, fmeasure=0.2424242424242424)

BERTScore - Precision: 0.9010605812072754, Recall: 0.6178848147392273, F1: 0.7330765724182129

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 1.0 indica que el 100% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.138 muestra que el sistema recupera solo el 13.8% de las palabras relevantes, lo que refleja un **recall extremadamente bajo**.
- **F1-Score:** 0.242 indica que, aunque el sistema es muy preciso, **recupera muy pocas palabras relevantes** en total.

2. ROUGE-2 Score:

- **Precision:** 0.857 indica que el 85.7% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.105 sugiere que el sistema recupera solo el 10.5% de los bigrams relevantes.
- **F1-Score:** 0.188 refleja que, aunque el sistema genera secuencias precisas, **recupera muy pocas secuencias relevantes** en general.

3. ROUGE-L Score:

- **Precision:** 1.0 sugiere que el sistema es completamente preciso en las secuencias más largas de coincidencias.
- **Recall:** 0.138 refleja que solo recupera una pequeña fracción de las secuencias relevantes.
- **F1-Score:** 0.242 muestra el **desequilibrio significativo** entre precisión y recall.

4. BERTScore:

- **Precision:** 0.901 indica que el 90.1% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.618 muestra que el sistema captura el 61.8% del contenido semántico relevante.
- **F1-Score:** 0.733 refleja un **buen rendimiento semántico con alta precisión**, aunque el recall podría mejorarse.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene una **precisión extremadamente alta** pero un **recall muy bajo**, lo que significa que genera respuestas muy precisas, pero solo recupera una pequeña parte del contenido relevante. El rendimiento en secuencias (ROUGE-2 y ROUGE-L) sigue esta misma tendencia, con un excelente desempeño en precisión pero una **baja capacidad de recuperación**. El **BERTScore** sugiere que el sistema es **semánticamente preciso**, pero el **recall semántico** también podría mejorarse. En general, el sistema es muy

eficiente en términos de precisión, pero necesita mejorar en **recuperación de información** para captar una mayor cantidad de contenido relevante.

Respuesta generada con RAG con Haystack:

Las operaciones USAR se dividen en cinco niveles, que son los siguientes:

1. ****Nivel 1: Evaluación general del área.****
2. ****Nivel 2: Evaluación del Triage del sitio de trabajo.****
3. ****Nivel 3: Búsqueda y rescate rápido.****
4. ****Nivel 4: Búsqueda y rescate total.****
5. ****Nivel 5: Búsqueda y recuperación con cobertura total.****

Cada nivel tiene definiciones y propósitos específicos y puede ser ejecutado en diferentes momentos y por diferentes equipos, dependiendo de las circunstancias y las necesidades de la operación.

Métricas:

```
ROUGE-1 score: Score(precision=0.6, recall=0.9310344827586207,
fmeasure=0.7297297297297297)
ROUGE-2 score: Score(precision=0.48314606741573035,
recall=0.7543859649122807, fmeasure=0.589041095890411)
ROUGE-L score: Score(precision=0.5555555555555556,
recall=0.8620689655172413, fmeasure=0.6756756756756757)
BERTScore - Precision: 0.8034967184066772, Recall: 0.9149895310401917, F1:
0.8556264042854309
```

Análisis de las Métricas:

1. ROUGE-1 Score:

- **Precision:** 0.6 indica que el 60% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.931 sugiere que el sistema recupera el 93.1% de las palabras relevantes, lo que refleja un **recall muy alto**.
- **F1-Score:** 0.730 muestra un **buen equilibrio** entre precisión y recall, lo que sugiere que el sistema es eficiente en la generación de contenido relevante con una buena capacidad de recuperación.

2. ROUGE-2 Score:

- **Precision:** 0.483 indica que el 48.3% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.754 muestra que el sistema recupera el 75.4% de los bigrams relevantes.
- **F1-Score:** 0.589 refleja un **rendimiento sólido** en la captura de secuencias significativas, con una **alta capacidad de recuperación** y una **precisión aceptable**.

3. ROUGE-L Score:

- **Precision:** 0.556 sugiere que el 55.6% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.

- **Recall:** 0.862 muestra que el sistema recupera el 86.2% de la secuencia relevante más larga.
 - **F1-Score:** 0.676 refleja un **buen equilibrio** entre precisión y recall en la generación de secuencias estructuradas.
4. **BERTScore:**
- **Precision:** 0.803 indica que el 80.3% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.915 sugiere que el sistema captura el 91.5% del contenido semántico relevante.
 - **F1-Score:** 0.856 refleja un **rendimiento muy sólido** en términos de correspondencia semántica, con un **buen equilibrio** entre precisión y recall.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **alto recall** y una **precisión moderada**, lo que sugiere que genera contenido relevante de manera eficiente, aunque con un margen para mejorar la precisión. El rendimiento en secuencias (ROUGE-2 y ROUGE-L) es **fuerte**, especialmente en la capacidad de recuperación de secuencias relevantes. El **BERTScore** muestra que el sistema tiene un **buen rendimiento semántico**, capturando la mayoría del contenido relevante de manera precisa y coherente. En general, el sistema es **muy eficiente en la recuperación de contenido relevante**, aunque mejorar la precisión podría hacer que las respuestas sean más concisas y exactas.

Respuesta generada con RAG con LlamaIndex

Las operaciones USAR se dividen en cinco niveles:

1. Nivel 1: Evaluación general del área.
2. Nivel 2: Evaluación del Triage del sitio de trabajo.
3. Nivel 3: Búsqueda y rescate rápido.
4. Nivel 4: Búsqueda y rescate total.
5. Nivel 5: Búsqueda y recuperación con cobertura total.

Métricas

```
ROUGE-1 score: Score(precision=0.9090909090909091,
recall=0.8620689655172413, fmeasure=0.8849557522123893)
ROUGE-2 score: Score(precision=0.7962962962962963,
recall=0.7543859649122807, fmeasure=0.7747747747747747)
ROUGE-L score: Score(precision=0.9090909090909091,
recall=0.8620689655172413, fmeasure=0.8849557522123893)
BERTScore - Precision: 0.9438149333000183, Recall: 0.9356445074081421, F1:
0.9397119879722595
```

Análisis de las Métricas

1. **ROUGE-1 Score:**
 - **Precision:** 0.909 indica que el 90.9% de las palabras generadas son relevantes en comparación con el texto de referencia.

- **Recall:** 0.862 sugiere que el sistema recupera el 86.2% de las palabras relevantes, lo que refleja un **recall muy alto**.
- **F1-Score:** 0.885 muestra un **excelente equilibrio** entre precisión y recall, lo que sugiere que el sistema genera contenido relevante de manera muy eficiente.
- 2. **ROUGE-2 Score:**
 - **Precision:** 0.796 indica que el 79.6% de los bigrams generados coinciden con el texto de referencia.
 - **Recall:** 0.754 muestra que el sistema recupera el 75.4% de los bigrams relevantes.
 - **F1-Score:** 0.775 refleja un **buen rendimiento** en la generación de secuencias significativas, con una **alta precisión y buena recuperación de secuencias**.
- 3. **ROUGE-L Score:**
 - **Precision:** 0.909 sugiere que el 90.9% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.862 muestra que el sistema recupera el 86.2% de la secuencia relevante más larga.
 - **F1-Score:** 0.885 refleja un **rendimiento sobresaliente** en la captura de secuencias estructuradas, con una **excelente precisión y recall alto**.
- 4. **BERTScore:**
 - **Precision:** 0.944 indica que el 94.4% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.936 sugiere que el sistema captura el 93.6% del contenido semántico relevante.
 - **F1-Score:** 0.940 refleja un **rendimiento semántico excepcional**, con un **equilibrio casi perfecto** entre precisión y recall en términos de significado.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican que el sistema tiene un **rendimiento excelente**, tanto en términos de precisión como de recall. El sistema es **muy eficiente en la recuperación de contenido relevante** y genera secuencias coherentes con **alta precisión y recuperación semántica excelente**. En general, el sistema es **altamente eficaz** y equilibrado, tanto en la literalidad como en la semántica de las respuestas, lo que lo convierte en una solución muy sólida para generar contenido relevante y preciso.

Pregunta 3: En qué consiste el nivel 1 referido a la Evaluación General del área?

Respuesta generada por LECHAT

Search...

Today


Como se compone el código d...

Today

Como se compone el código d...

CP

En qué consiste el nivel 1 referido a la Evaluación General del área?



En el contexto de las operaciones de Urban Search and Rescue (USAR), el Nivel 1 de la Evaluación General del Área se refiere a la fase inicial de la evaluación y respuesta a una emergencia. Esta fase es crucial para obtener una comprensión rápida y precisa de la situación, lo que permite a los equipos de rescate tomar decisiones informadas y coordinar sus esfuerzos de manera efectiva. A continuación se describen los componentes clave del Nivel 1 de la Evaluación General del Área:

1. Reconocimiento Inicial:

- **Evaluación Visual:** Los primeros en responder realizan una evaluación visual rápida del área afectada para identificar peligros inmediatos, como estructuras inestables, fugas de gas, líneas eléctricas caídas, etc.
- **Identificación de Víctimas:** Se busca identificar a las víctimas que requieren atención inmediata, ya sea por lesiones graves o por estar atrapadas.

2. Evaluación de Peligros:

- **Peligros Inmediatos:** Se identifican y evalúan los peligros inmediatos que podrían poner en riesgo a los rescatistas y a las víctimas, como incendios, inundaciones, derrumbes, etc.
- **Peligros Secundarios:** Se consideran los peligros secundarios que podrían surgir como resultado del incidente, como réplicas de terremotos, deslizamientos de tierra, etc.

3. Establecimiento de un Perímetro de Seguridad:

- **Área Segura:** Se establece un perímetro de seguridad alrededor del área afectada para proteger a los rescatistas y a las víctimas de peligros adicionales.
- **Control de Acceso:** Se controla el acceso al área para asegurar que solo el personal autorizado y equipado adecuadamente pueda entrar.

4. Comunicación y Coordinación:

- **Informe Inicial:** Se proporciona un informe inicial a los coordinadores de la respuesta, detallando la situación, los peligros identificados y las necesidades inmediatas.
- **Coordinación con Otros Equipos:** Se coordina con otros equipos de rescate, servicios de emergencia y autoridades locales para asegurar una respuesta integrada y eficiente.

5. Evaluación de Recursos:

- **Recursos Disponibles:** Se evalúan los recursos disponibles en el área, como equipos de rescate, suministros médicos, vehículos de emergencia, etc.
- **Necesidades Adicionales:** Se identifican las necesidades adicionales de recursos y se solicitan refuerzos si es necesario.

6. Planificación Inicial:

- **Prioridades de Rescate:** Se establecen las prioridades de rescate basadas en la evaluación inicial, dando prioridad a las víctimas con lesiones graves y a aquellas que están atrapadas.
- **Asignación de Tareas:** Se asignan tareas específicas a los equipos de rescate, como la búsqueda y rescate, la estabilización de estructuras, la atención médica, etc.

El Nivel 1 de la Evaluación General del Área es esencial para establecer una base sólida para la respuesta a la emergencia. Permite a los equipos de rescate actuar de manera rápida y efectiva, minimizando los riesgos y maximizando las posibilidades de salvar vidas.

Transcripción.

En el contexto de las operaciones de Urban Search and Rescue (USAR), el Nivel 1 de la Evaluación General del Área se refiere a la fase inicial de la evaluación y respuesta a una emergencia. Esta fase es crucial para obtener una comprensión rápida y precisa de la situación, lo que permite a los equipos de rescate tomar decisiones informadas y coordinar sus esfuerzos de manera efectiva. A continuación se describen los componentes clave del Nivel 1 de la Evaluación General del Área:

1. Reconocimiento Inicial:

- Evaluación Visual: Los primeros en responder realizan una evaluación visual rápida del área afectada para identificar peligros inmediatos, como estructuras inestables, fugas de gas, líneas eléctricas caídas, etc.
 - Identificación de Víctimas: Se busca identificar a las víctimas que requieren atención inmediata, ya sea por lesiones graves o por estar atrapadas.
2. Evaluación de Peligros:
- Peligros Inmediatos: Se identifican y evalúan los peligros inmediatos que podrían poner en riesgo a los rescatistas y a las víctimas, como incendios, inundaciones, derrumbes, etc.
 - Peligros Secundarios: Se consideran los peligros secundarios que podrían surgir como resultado del incidente, como réplicas de terremotos, deslizamientos de tierra, etc.
3. Establecimiento de un Perímetro de Seguridad:
- Área Segura: Se establece un perímetro de seguridad alrededor del área afectada para proteger a los rescatistas y a las víctimas de peligros adicionales.
 - Control de Acceso: Se controla el acceso al área para asegurar que solo el personal autorizado y equipado adecuadamente pueda entrar.
4. Comunicación y Coordinación:
- Informe Inicial: Se proporciona un informe inicial a los coordinadores de la respuesta, detallando la situación, los peligros identificados y las necesidades inmediatas.
 - Coordinación con Otros Equipos: Se coordina con otros equipos de rescate, servicios de emergencia y autoridades locales para asegurar una respuesta integrada y eficiente.
5. Evaluación de Recursos:
- Recursos Disponibles: Se evalúan los recursos disponibles en el área, como equipos de rescate, suministros médicos, vehículos de emergencia, etc.
 - Necesidades Adicionales: Se identifican las necesidades adicionales de recursos y se solicitan refuerzos si es necesario.
6. Planificación Inicial:
- Prioridades de Rescate: Se establecen las prioridades de rescate basadas en la evaluación inicial, dando prioridad a las víctimas con lesiones graves y a aquellas que están atrapadas.
 - Asignación de Tareas: Se asignan tareas específicas a los equipos de rescate, como la búsqueda y rescate, la estabilización de estructuras, la atención médica, etc.

El Nivel 1 de la Evaluación General del Área es esencial para establecer una base sólida para la respuesta a la emergencia. Permite a los equipos de rescate actuar de manera rápida y efectiva, minimizando los riesgos y maximizando las posibilidades de salvar vidas.

Métricas sin RAG.

ROUGE-1 score: Score(precision=0.16527196652719664, recall=0.5984848484848485, fmeasure=0.2590163934426229)
ROUGE-2 score: Score(precision=0.0440251572327044, recall=0.16030534351145037, fmeasure=0.06907894736842106)
ROUGE-L score: Score(precision=0.08786610878661087, recall=0.3181818181818182, fmeasure=0.13770491803278687)
BERTScore - Precision: 0.6448379755020142, Recall: 0.694067120552063, F1: 0.668547511100769

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.165 indica que solo el 16.5% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.598 sugiere que el sistema recupera el 59.8% de las palabras relevantes, lo que refleja un **recall moderado** pero con una **precisión baja**.
- **F1-Score:** 0.259 muestra un **desequilibrio** significativo entre precisión y recall, lo que sugiere que el sistema recupera contenido relevante, pero también genera una cantidad considerable de contenido irrelevante.

2. ROUGE-2 Score:

- **Precision:** 0.044 indica que solo el 4.4% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.160 muestra que el sistema recupera el 16% de los bigrams relevantes.
- **F1-Score:** 0.069 refleja un **rendimiento bajo** en la captura de secuencias de palabras significativas, con una **precisión extremadamente baja**.

3. ROUGE-L Score:

- **Precision:** 0.088 sugiere que solo el 8.8% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.318 indica que el sistema recupera el 31.8% de la secuencia relevante más larga.
- **F1-Score:** 0.138 refleja un **rendimiento bajo** en la captura de secuencias estructuradas, con **baja precisión** y un **recall limitado**.

4. BERTScore:

- **Precision:** 0.645 indica que el 64.5% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.694 muestra que el sistema captura el 69.4% del contenido semántico relevante.
- **F1-Score:** 0.669 refleja un **rendimiento semántico moderado**, con un **equilibrio aceptable** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **recall moderado** pero una **precisión baja**, lo que sugiere que genera una cantidad considerable de contenido irrelevante. El rendimiento en secuencias (ROUGE-2 y ROUGE-L) es especialmente bajo, lo que indica dificultades en generar relaciones significativas entre palabras y secuencias coherentes. El **BERTScore** muestra un **rendimiento semántico moderado**, lo que sugiere que el sistema

captura de manera razonable el significado general, pero necesita mejorar en **precisión literal** y en la **estructura de las secuencias** para generar respuestas más concisas y relevantes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

El reconocimiento preliminar del área afectada o asignada.

Con el objetivo de:

Determinar el alcance y la magnitud del incidente.

Identificación del alcance, ubicación y tipos de daños.

Estimativo de las necesidades urgentes de rescate.

Desarrollo del plan de sectorización.

Establecimiento de prioridades.

Identificación de riesgos generales.

Identificación de temas estructurales.

Identificación de ubicaciones potenciales para las Bases de Operaciones.

Usualmente realizado por medio de vehículos, helicópteros, naves acuáticas, a pie o por medio de reportes de otros, por ejemplo, de LEMA.

Rápida verificación visual inicial del área asignada o afectada.

Los equipos que ejecuten este nivel de tareas deben permanecer móviles y no involucrarse en operaciones de rescate y reportar los resultados tan rápidamente como sea posible.

Respuesta generada con RAG Básico:

El nivel 1, referido a la Evaluación General del área, consiste en el reconocimiento preliminar del área afectada o asignada. Los objetivos de este nivel incluyen:

- Determinar el alcance y la magnitud del incidente.
- Identificar el alcance, ubicación y tipos de daños.
- Estimar las necesidades urgentes de rescate.
- Desarrollar un plan de sectorización.
- Establecer prioridades.
- Identificar riesgos generales.
- Identificar temas estructurales.
- Identificar ubicaciones potenciales para las Bases de Operaciones.

Esta evaluación es usualmente realizada por medio de vehículos, helicópteros, naves acuáticas, a pie o por medio de reportes de otros, como los de LEMA. Los equipos que ejecuten este nivel de tareas deben permanecer

móviles y no involucrarse en operaciones de rescate, reportando los resultados tan rápidamente como sea posible.

Métricas:

```
ROUGE-1 score: Score(precision=0.7769230769230769,  
recall=0.7651515151515151, fmeasure=0.7709923664122137)  
ROUGE-2 score: Score(precision=0.6356589147286822,  
recall=0.6259541984732825, fmeasure=0.6307692307692307)  
ROUGE-L score: Score(precision=0.7384615384615385,  
recall=0.7272727272727273, fmeasure=0.732824427480916)  
BERTScore - Precision: 0.8849087357521057, Recall: 0.8986932039260864, F1:  
0.8917476534843445
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.777 indica que el 77.7% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.765 sugiere que el sistema recupera el 76.5% de las palabras relevantes, lo que refleja un **buen equilibrio** entre precisión y recall.
- **F1-Score:** 0.771 muestra un **rendimiento sólido**, lo que sugiere que el sistema genera una gran cantidad de contenido relevante de manera eficiente.

2. ROUGE-2 Score:

- **Precision:** 0.636 indica que el 63.6% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.626 muestra que el sistema recupera el 62.6% de los bigrams relevantes.
- **F1-Score:** 0.631 refleja un **buen rendimiento** en la captura de secuencias significativas, con un equilibrio aceptable entre precisión y recall.

3. ROUGE-L Score:

- **Precision:** 0.738 indica que el 73.8% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.727 sugiere que el sistema recupera el 72.7% de la secuencia relevante más larga.
- **F1-Score:** 0.733 refleja un **rendimiento sólido** en la generación de secuencias estructuradas.

4. BERTScore:

- **Precision:** 0.885 indica que el 88.5% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.899 sugiere que el sistema captura el 89.9% del contenido semántico relevante.
- **F1-Score:** 0.892 refleja un **rendimiento excelente** en términos de correspondencia semántica, con un **muy buen equilibrio** entre precisión y recall.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican que el sistema tiene un **rendimiento fuerte** tanto en términos de precisión como de recall. El sistema genera contenido **relevante y**

estructurado con alta precisión y recupera una cantidad significativa de secuencias y palabras clave. El **BERTScore** refuerza que el sistema es **muy eficaz en la captura del significado semántico**, con un excelente equilibrio entre precisión y recall. En general, el sistema es **muy eficiente y equilibrado**, generando respuestas relevantes y bien estructuradas.

Respuesta generada con RAG con Haystack:

El nivel 1 referido a la Evaluación General del área consiste en el reconocimiento preliminar del área afectada o asignada. El objetivo de esta evaluación es determinar el alcance y la magnitud del incidente, identificar el alcance, ubicación y tipos de daños, estimar las necesidades urgentes de rescate, desarrollar un plan de sectorización, establecer prioridades, identificar riesgos generales y estructurales, y determinar ubicaciones potenciales para las Bases de Operaciones. Esta evaluación se realiza usualmente por medio de vehículos, helicópteros, naves acuáticas, a pie o por medio de reportes de otros, como LEMA. Los equipos que ejecuten este nivel de tareas deben permanecer móviles y no involucrarse en operaciones de rescate, reportando los resultados tan rápidamente como sea posible.

Métricas:

```
ROUGE-1 score: Score(precision=0.7829457364341085,
recall=0.7651515151515151, fmeasure=0.7739463601532567)
ROUGE-2 score: Score(precision=0.6328125, recall=0.6183206106870229,
fmeasure=0.6254826254826255)
ROUGE-L score: Score(precision=0.7364341085271318,
recall=0.7196969696969697, fmeasure=0.7279693486590039)
BERTScore - Precision: 0.8877774477005005, Recall: 0.8814699649810791, F1:
0.8846124410629272
```

Análisis de las Métricas.

1. ROUGE-1 Score:

- **Precision:** 0.783 indica que el 78.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.765 sugiere que el sistema recupera el 76.5% de las palabras relevantes, lo que refleja un **equilibrio sólido** entre precisión y recall.
- **F1-Score:** 0.774 muestra un **rendimiento fuerte**, indicando que el sistema genera una gran cantidad de contenido relevante de manera eficiente.

2. ROUGE-2 Score:

- **Precision:** 0.633 indica que el 63.3% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.618 muestra que el sistema recupera el 61.8% de los bigrams relevantes.
- **F1-Score:** 0.625 refleja un **buen rendimiento** en la captura de secuencias significativas, con un equilibrio aceptable entre precisión y recall.

3. ROUGE-L Score:

- **Precision:** 0.736 indica que el 73.6% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.720 sugiere que el sistema recupera el 72.0% de la secuencia relevante más larga.

- **F1-Score:** 0.728 refleja un **rendimiento sólido** en la generación de secuencias estructuradas.
- 4. **BERTScore:**
 - **Precision:** 0.888 indica que el 88.8% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.881 muestra que el sistema captura el 88.1% del contenido semántico relevante.
 - **F1-Score:** 0.885 refleja un **rendimiento excelente** en términos de correspondencia semántica, con un **muy buen equilibrio** entre precisión y recall.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican que el sistema tiene un **rendimiento sólido y equilibrado** en términos de precisión y recall, tanto en la generación de palabras individuales como en la captura de secuencias significativas. El **BERTScore** refuerza que el sistema es **altamente eficiente en la captura del significado semántico**, con un equilibrio casi perfecto entre precisión y recall. En general, el sistema es **muy eficiente y bien estructurado**, generando respuestas relevantes tanto en literalidad como en semántica.

Respuesta generada con RAG con LLaIndex

El nivel 1 de la Evaluación General del área consiste en el reconocimiento preliminar del área afectada o asignada. Su objetivo es determinar el alcance y la magnitud del incidente, identificar el alcance, ubicación y tipos de daños, estimar las necesidades urgentes de rescate, desarrollar un plan de sectorización, establecer prioridades, identificar riesgos generales y temas estructurales, y determinar ubicaciones potenciales para las Bases de Operaciones. Esta evaluación se realiza usualmente mediante vehículos, helicópteros, naves acuáticas, a pie o a través de reportes de otros, como LEMA. Los equipos que ejecuten este nivel de tareas deben permanecer móviles y no involucrarse en operaciones de rescate, reportando los resultados tan rápidamente como sea posible.

Métricas

ROUGE-1 score: Score(precision=0.7741935483870968,
recall=0.7272727272727273, fmeasure=0.7500000000000001)
ROUGE-2 score: Score(precision=0.6016260162601627,
recall=0.5648854961832062, fmeasure=0.5826771653543307)
ROUGE-L score: Score(precision=0.717741935483871,
recall=0.6742424242424242, fmeasure=0.6953125)
BERTScore - Precision: 0.8914587497711182, Recall: 0.876801609992981, F1:
0.8840694427490234

Análisis de las Métricas

1. **ROUGE-1 Score:**
 - **Precision:** 0.774 indica que el 77.4% de las palabras generadas son relevantes en comparación con el texto de referencia.
 - **Recall:** 0.727 sugiere que el sistema recupera el 72.7% de las palabras relevantes.

- **F1-Score:** 0.750 refleja un **buen equilibrio** entre precisión y recall, lo que indica que el sistema genera una cantidad significativa de contenido relevante de manera eficiente.
- 2. **ROUGE-2 Score:**
 - **Precision:** 0.602 indica que el 60.2% de los bigrams generados coinciden con el texto de referencia.
 - **Recall:** 0.565 muestra que el sistema recupera el 56.5% de los bigrams relevantes.
 - **F1-Score:** 0.583 refleja un **buen rendimiento** en la captura de secuencias de dos palabras, con una **precisión razonable** y una **recuperación sólida**.
- 3. **ROUGE-L Score:**
 - **Precision:** 0.718 sugiere que el 71.8% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.674 indica que el sistema recupera el 67.4% de la secuencia relevante más larga.
 - **F1-Score:** 0.695 muestra un **buen rendimiento** en la captura de secuencias estructuradas, con un **equilibrio aceptable** entre precisión y recall.
- 4. **BERTScore:**
 - **Precision:** 0.891 indica que el 89.1% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.877 sugiere que el sistema captura el 87.7% del contenido semántico relevante.
 - **F1-Score:** 0.884 refleja un **rendimiento semántico excelente**, con un **equilibrio fuerte** entre precisión y recall.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **rendimiento sólido** en términos de precisión y recall, tanto en la generación de palabras individuales como en la captura de secuencias. El **BERTScore** refuerza la idea de que el sistema es **muy eficiente en la captura del significado semántico**, mostrando un **alto equilibrio** entre precisión y recall. En general, el sistema es **robusto** en la generación de respuestas, tanto en términos de literalidad como de semántica, con una buena capacidad de capturar contenido relevante y coherente.

Pregunta 4: En el nivel 2 referido a la Evaluación del triage del sitio de trabajo, ¿cuáles son las herramientas INSARAG?

Respuesta generada por LECHAT

Search...

Today

Como se compone el código d...

CP

En el nivel 2 referido a la Evaluación del triage del sitio de trabajo, ¿cuáles son las herramientas INSARAG?

IM

En el contexto de las operaciones de Urban Search and Rescue (USAR), el Nivel 2 de la Evaluación del Triage del Sitio de Trabajo se refiere a la fase en la que se realiza una evaluación más detallada y sistemática del área afectada para priorizar las acciones de rescate y atención médica. El Grupo Asesor Internacional de Búsqueda y Rescate (INSARAG, por sus siglas en inglés) ha desarrollado una serie de herramientas y métodos para facilitar esta evaluación. A continuación se describen algunas de las herramientas INSARAG utilizadas en el Nivel 2:

1. INSARAG Guidelines:

Manual de Campo INSARAG: Este manual proporciona directrices detalladas sobre cómo realizar evaluaciones de triage en el sitio de trabajo, incluyendo procedimientos para la evaluación de estructuras, la búsqueda y rescate de víctimas, y la atención médica inicial.

Listas de Verificación: INSARAG proporciona listas de verificación que ayudan a los equipos de rescate a asegurarse de que todos los aspectos importantes de la evaluación del triage sean considerados.

2. INSARAG Triage Tags:

Etiquetas de Triage: Estas etiquetas se utilizan para clasificar a las víctimas según la gravedad de sus lesiones y la urgencia de la atención médica. Las etiquetas suelen estar codificadas por colores, donde cada color representa una categoría de prioridad (por ejemplo, rojo para víctimas críticas, amarillo para víctimas con lesiones graves pero no inmediatamente mortales, verde para víctimas con lesiones menores, y negro para víctimas fallecidas).

3. INSARAG Structural Assessment Forms:

Formularios de Evaluación Estructural: Estos formularios ayudan a los equipos de rescate a evaluar la estabilidad de las estructuras dañadas y a determinar si es seguro entrar en ellas para realizar operaciones de búsqueda y rescate.

4. INSARAG Medical Assessment Tools:

Herramientas de Evaluación Médica: Incluyen formularios y listas de verificación que ayudan a los equipos médicos a realizar evaluaciones rápidas y precisas de las víctimas, identificando lesiones y determinando la prioridad de atención.

5. INSARAG Incident Command System (ICS):

Sistema de Comando de Incidentes: INSARAG utiliza el ICS para estructurar y coordinar la respuesta a la emergencia. Este sistema incluye roles y responsabilidades claras para los diferentes miembros del equipo, así como procedimientos para la comunicación y la toma de decisiones.

6. INSARAG Mapping and Documentation Tools:

Herramientas de Mapeo y Documentación: Incluyen software y equipos para mapear el área afectada y documentar la ubicación de las víctimas, los peligros y los recursos disponibles. Estas herramientas son esenciales para mantener una visión clara y actualizada de la situación en el sitio de trabajo.

7. INSARAG Communication Equipment:

Equipos de Comunicación: INSARAG proporciona directrices sobre el uso de equipos de comunicación, como radios, sistemas de satélite y dispositivos de comunicación portátiles, para asegurar una comunicación efectiva entre los equipos de rescate y los coordinadores de la respuesta.

Estas herramientas y métodos de INSARAG son fundamentales para realizar una evaluación del triage del sitio de trabajo de manera eficiente y efectiva. Permiten a los equipos de rescate priorizar sus acciones, minimizar los riesgos y maximizar las posibilidades de salvar vidas en situaciones de emergencia.

Transcripción.

En el contexto de las operaciones de Urban Search and Rescue (USAR), el Nivel 2 de la Evaluación del Triage del Sitio de Trabajo se refiere a la fase en la que se realiza una evaluación más detallada y sistemática del área afectada para priorizar las acciones de rescate y atención médica. El Grupo Asesor Internacional de Búsqueda y Rescate (INSARAG, por sus siglas en inglés) ha desarrollado una serie de herramientas y métodos para facilitar esta evaluación. A continuación se describen algunas de las herramientas INSARAG utilizadas en el Nivel 2:

1. INSARAG Guidelines:

- Manual de Campo INSARAG: Este manual proporciona directrices detalladas sobre cómo realizar evaluaciones de triage en el sitio de trabajo, incluyendo procedimientos para la evaluación de estructuras, la búsqueda y rescate de víctimas, y la atención médica inicial.

- Listas de Verificación: INSARAG proporciona listas de verificación que ayudan a los equipos de rescate a asegurarse de que todos los aspectos importantes de la evaluación del triage sean considerados.

2. INSARAG Triage Tags:

- Etiquetas de Triage: Estas etiquetas se utilizan para clasificar a las víctimas según la gravedad de sus lesiones y la urgencia de la atención médica. Las etiquetas suelen estar codificadas por colores, donde cada color representa una categoría de prioridad (por ejemplo, rojo para víctimas críticas, amarillo para víctimas con lesiones graves pero no inmediatamente mortales, verde para víctimas con lesiones menores, y negro para víctimas fallecidas).

3. INSARAG Structural Assessment Forms:

- Formularios de Evaluación Estructural: Estos formularios ayudan a los equipos de rescate a evaluar la estabilidad de las estructuras dañadas y a determinar si es seguro entrar en ellas para realizar operaciones de búsqueda y rescate.

4. INSARAG Medical Assessment Tools:

- Herramientas de Evaluación Médica: Incluyen formularios y listas de verificación que ayudan a los equipos médicos a realizar evaluaciones rápidas y precisas de las víctimas, identificando lesiones y determinando la prioridad de atención.

5. INSARAG Incident Command System (ICS):

- Sistema de Comando de Incidentes: INSARAG utiliza el ICS para estructurar y coordinar la respuesta a la emergencia. Este sistema incluye roles y responsabilidades claras para los diferentes miembros del equipo, así como procedimientos para la comunicación y la toma de decisiones.

6. INSARAG Mapping and Documentation Tools:

- Herramientas de Mapeo y Documentación: Incluyen software y equipos para mapear el área afectada y documentar la ubicación de las víctimas, los peligros y los recursos disponibles. Estas herramientas son esenciales para mantener una visión clara y actualizada de la situación en el sitio de trabajo.

7. INSARAG Communication Equipment:

- Equipos de Comunicación: INSARAG proporciona directrices sobre el uso de equipos de comunicación, como radios, sistemas de satélite y dispositivos de comunicación portátiles, para asegurar una comunicación efectiva entre los equipos de rescate y los coordinadores de la respuesta.

Estas herramientas y métodos de INSARAG son fundamentales para realizar una evaluación del triage del sitio de trabajo de manera eficiente y efectiva. Permiten a los equipos de rescate priorizar sus acciones, minimizar los riesgos y maximizar las posibilidades de salvar vidas en situaciones de emergencia.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.052158273381294966, recall=0.5,
fmeasure=0.09446254071661238)
ROUGE-2 score: Score(precision=0.016216216216216217,
recall=0.15789473684210525, fmeasure=0.029411764705882356)
ROUGE-L score: Score(precision=0.04136690647482014,
recall=0.39655172413793105, fmeasure=0.0749185667752443)
BERTScore - Precision: 0.6177817583084106, Recall: 0.7023574113845825, F1:
0.6573603749275208
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.052 indica que solo el 5.2% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.500 muestra que el sistema recupera el 50% de las palabras relevantes, lo que indica un **recall moderado** pero con una **precisión muy baja**.
- **F1-Score:** 0.094 refleja un **desequilibrio significativo** entre precisión y recall, lo que sugiere que el sistema genera mucho contenido irrelevante.

2. ROUGE-2 Score:

- **Precision:** 0.016 indica que solo el 1.6% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.158 sugiere que el sistema recupera el 15.8% de los bigrams relevantes.
- **F1-Score:** 0.029 refleja un **rendimiento muy bajo** en la captura de secuencias de palabras, con una **precisión extremadamente baja**.

3. ROUGE-L Score:

- **Precision:** 0.041 sugiere que solo el 4.1% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.397 indica que el sistema recupera el 39.7% de la secuencia relevante más larga.
- **F1-Score:** 0.075 refleja un **rendimiento bajo** en la captura de secuencias estructuradas, con **baja precisión** y un **recall moderado**.

4. BERTScore:

- **Precision:** 0.618 indica que el 61.8% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.702 muestra que el sistema captura el 70.2% del contenido semántico relevante.
- **F1-Score:** 0.657 refleja un **rendimiento semántico moderado**, con un equilibrio aceptable entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **recall moderado**, pero una **precisión extremadamente baja**, lo que sugiere que genera una cantidad significativa de

contenido irrelevante. El rendimiento en secuencias (ROUGE-2 y ROUGE-L) es particularmente bajo, lo que indica dificultades en generar relaciones significativas entre palabras y secuencias coherentes. El **BERTScore** muestra un **rendimiento semántico razonable**, lo que sugiere que el sistema capta bien el significado general, aunque necesita mejorar considerablemente en **precisión literal y estructura de secuencias** para generar respuestas más relevantes y concisas.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

1. Formato del Triage del Sitio de Trabajo. 2. Resumen de UCC.
3. Un mapa del área del sector siendo evaluada es altamente recomendado y debe ser usado para tener claras las áreas evaluadas y despejadas. También puede haber información como: resúmenes de LEMA, información de los equipos locales, coordenadas GPS, fotografías, etc.

Respuesta generada con RAG Básico:

En el nivel 2 referido a la Evaluación del triage del sitio de trabajo, las herramientas INSARAG son:

1. Formato del Triage del Sitio de Trabajo.
2. Resumen de UCC.
3. Un mapa del área del sector siendo evaluada es altamente recomendado y debe ser usado para tener claras las áreas evaluadas y despejadas. También puede haber información como: resúmenes de LEMA, información de los equipos locales, coordenadas GPS, fotografías, etc.

Métricas:

ROUGE-1 score: Score(precision=0.7532467532467533, recall=1.0, fmeasure=0.8592592592592593)
ROUGE-2 score: Score(precision=0.75, recall=1.0, fmeasure=0.8571428571428571)
ROUGE-L score: Score(precision=0.7532467532467533, recall=1.0, fmeasure=0.8592592592592593)
BERTScore - Precision: 0.8967797756195068, Recall: 0.9635167121887207, F1: 0.9289512038230896

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precision:** 0.753 indica que el 75.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 1.0 muestra que el sistema recupera el 100% de las palabras relevantes, lo que refleja un **recall perfecto**.
- **F1-Score:** 0.859 indica un **equilibrio sólido** entre precisión y recall, lo que sugiere que el sistema genera una cantidad significativa de contenido relevante con una **recuperación completa**.

2. **ROUGE-2 Score:**

- **Precision:** 0.75 indica que el 75% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 1.0 sugiere que el sistema recupera el 100% de los bigrams relevantes.
- **F1-Score:** 0.857 refleja un **rendimiento excelente** en la captura de secuencias de palabras, con una **recuperación perfecta** y una **alta precisión**.

3. **ROUGE-L Score:**

- **Precision:** 0.753 sugiere que el 75.3% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 1.0 muestra que el sistema recupera la secuencia relevante más larga en su totalidad.
- **F1-Score:** 0.859 refleja un **rendimiento fuerte** en la captura de secuencias estructuradas.

4. **BERTScore:**

- **Precision:** 0.897 indica que el 89.7% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.964 sugiere que el sistema captura el 96.4% del contenido semántico relevante.
- **F1-Score:** 0.929 refleja un **rendimiento semántico excepcional**, con un **equilibrio fuerte** entre precisión y recall.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican que el sistema tiene un **rendimiento sobresaliente** tanto en términos de precisión como de recall. El sistema es **extremadamente eficaz en la recuperación de contenido relevante**, generando secuencias de palabras y contenido con **altísima precisión** y **recall perfecto**. El **BERTScore** refuerza que el sistema es **altamente preciso en términos semánticos**, capturando el significado de manera casi perfecta. En general, el sistema es **muy robusto** y eficiente en la generación de contenido relevante y bien estructurado.

Respuesta generada con RAG con Haystack:

En el nivel 2 referido a la Evaluación del triage del sitio de trabajo, las herramientas INSARAG son:

1. Formato del Triage del Sitio de Trabajo.
2. Resumen de UCC.
3. Un mapa del área del sector siendo evaluada es altamente recomendado y debe ser usado para tener claras las áreas evaluadas y despejadas. También puede haber información como: resúmenes de LEMA, información de los equipos locales, coordenadas GPS, fotografías, etc.

Métricas:

ROUGE-1 score: Score(precision=0.7532467532467533, recall=1.0, fmeasure=0.8592592592592593)
ROUGE-2 score: Score(precision=0.75, recall=1.0, fmeasure=0.8571428571428571)

ROUGE-L score: Score(precision=0.7532467532467533, recall=1.0, fmeasure=0.8592592592592593)
BERTScore - Precision: 0.8967797756195068, Recall: 0.9635167121887207, F1: 0.9289512038230896

Análisis de las Métricas.

1. ROUGE-1:

- **Precisión:** 75.3% de las palabras generadas son relevantes.
- **Recall:** 100% de las palabras relevantes fueron recuperadas.
- **F1-Score:** 0.859 refleja un **balance fuerte** entre precisión y recall, lo que significa que el sistema genera la mayoría del contenido relevante.

2. ROUGE-2:

- **Precisión:** 75% de los bigrams generados son correctos.
- **Recall:** 100% de los bigrams relevantes fueron recuperados.
- **F1-Score:** 0.857 indica una **alta precisión y recuperación completa** de secuencias significativas.

3. ROUGE-L:

- **Precisión y recall:** Ambas en 75.3% y 100%, respectivamente, lo que refuerza el **alto rendimiento estructural** del sistema.
- **F1-Score:** 0.859 muestra una **recuperación estructural eficiente**.

4. BERTScore:

- **Precisión:** 89.7% de las palabras generadas son semánticamente relevantes.
- **Recall:** 96.4% de las palabras relevantes fueron recuperadas.
- **F1-Score:** 0.929 refleja un **alto rendimiento semántico**.

Conclusión: El sistema presenta una **excelente precisión y recall perfecto**, lo que significa que genera respuestas muy relevantes, tanto en términos literales como semánticos.

Respuesta generada con RAG con LlamaIndex

Las herramientas INSARAG para el nivel 2, referido a la Evaluación del triage del sitio de trabajo, incluyen el Formato Triage del Sitio de Trabajo, el Formato de Reporte del Sitio de Trabajo, el Formato de Liberación de Víctima y el Sistema de Marcaje del Sitio de Trabajo.

Métricas

ROUGE-1 score: Score(precision=0.35294117647058826, recall=0.3103448275862069, fmeasure=0.3302752293577982)
ROUGE-2 score: Score(precision=0.12, recall=0.10526315789473684, fmeasure=0.11214953271028036)
ROUGE-L score: Score(precision=0.21568627450980393, recall=0.1896551724137931, fmeasure=0.20183486238532108)
BERTScore - Precision: 0.7349568009376526, Recall: 0.6849461197853088, F1: 0.7090707421302795

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.353 indica que el 35.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.310 muestra que el sistema recupera el 31.0% de las palabras relevantes.
- **F1-Score:** 0.330 refleja un **equilibrio moderado** entre precisión y recall, pero con espacio para mejorar ambos aspectos.

2. ROUGE-2 Score:

- **Precisión:** 0.120 indica que solo el 12.0% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.105 sugiere que el sistema recupera solo el 10.5% de los bigrams relevantes.
- **F1-Score:** 0.112 refleja un **rendimiento bajo** en la captura de secuencias de palabras significativas, con una **precisión limitada** y un **recall bajo**.

3. ROUGE-L Score:

- **Precisión:** 0.216 sugiere que solo el 21.6% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.190 indica que el sistema recupera el 19.0% de la secuencia relevante más larga.
- **F1-Score:** 0.202 refleja un **rendimiento bajo** en la generación de secuencias estructuradas, con espacio para mejorar tanto en precisión como en recall.

4. BERTScore:

- **Precisión:** 0.735 indica que el 73.5% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.685 sugiere que el sistema captura el 68.5% del contenido semántico relevante.
- **F1-Score:** 0.709 refleja un **rendimiento semántico razonablemente sólido**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene un **rendimiento moderado** en la generación de contenido relevante, pero la **precisión y el recall** son bajos, especialmente en la captura de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** muestra que el sistema es **semánticamente más fuerte**, capturando el significado general del texto con mayor eficiencia. Para mejorar, el sistema debe centrarse en aumentar la **precisión** y **recuperar secuencias más coherentes** para generar respuestas más relevantes.

Pregunta 5: Cuáles son las definiciones y propósito del nivel 4: Búsqueda y rescate total?

Respuesta generada por LECHAT

Search...

Today

Como se compone el código d...

Today

Como se compone el código d...

Today

Como se compone el código d...

CP

Cuáles son las definiciones y propósito del nivel 4: Búsqueda y rescate total?

M

El Nivel 4 en las operaciones de Urban Search and Rescue (USAR) se refiere a la fase de "Búsqueda y Rescate Total" (Full Search and Rescue). Esta fase es crucial y representa la culminación de los esfuerzos de evaluación y planificación previos. A continuación se describen las definiciones y el propósito del Nivel 4: Búsqueda y Rescate Total.

Definiciones

1. Búsqueda y Rescate Total:

Búsqueda: La acción de localizar a las víctimas atrapadas o desaparecidas en el área afectada. Esto incluye el uso de técnicas avanzadas de búsqueda, como el uso de perros de rescate, equipos de detección de sonido y cámaras térmicas.

Rescate: La acción de liberar y extraer a las víctimas de manera segura de las estructuras colapsadas o de otras situaciones peligrosas. Esto puede involucrar el uso de herramientas especializadas, como sierras, gatos hidráulicos y equipos de levantamiento.

Propósito

El propósito del Nivel 4: Búsqueda y Rescate Total es maximizar las posibilidades de encontrar y rescatar a todas las víctimas atrapadas o desaparecidas en el área afectada. Esta fase se caracteriza por una serie de actividades y objetivos específicos:

1. Localización de Víctimas:

Técnicas Avanzadas: Utilizar técnicas avanzadas de búsqueda para localizar a las víctimas, incluyendo el uso de tecnología y equipos especializados.

Cobertura Completa: Asegurarse de que todas las áreas del sitio de trabajo sean revisadas exhaustivamente para no dejar a ninguna víctima sin ser encontrada.

2. Extracción Segura:

Estabilización de Estructuras: Estabilizar las estructuras dañadas para asegurar que los rescatistas puedan trabajar de manera segura.

Uso de Herramientas Especializadas: Utilizar herramientas y equipos especializados para liberar y extraer a las víctimas de manera segura y eficiente.

3. Atención Médica Inmediata:

Evaluación Médica: Evaluar el estado médico de las víctimas rescatadas y proporcionar atención médica inmediata según sea necesario.

Transporte a Instalaciones Médicas: Coordinar el transporte de las víctimas a instalaciones médicas para recibir atención adicional.

4. Coordinación y Comunicación:

Coordinación de Equipos: Coordinar los esfuerzos de los diferentes equipos de rescate para asegurar una respuesta integrada y eficiente.

Comunicación Continua: Mantener una comunicación continua entre los equipos de rescate, los coordinadores de la respuesta y otras agencias involucradas.

5. Documentación y Registro:

Registro de Víctimas: Documentar y registrar a todas las víctimas encontradas y rescatadas para mantener un registro preciso de la situación.

Informes de Progreso: Proporcionar informes regulares sobre el progreso de las operaciones de búsqueda y rescate a los coordinadores de la respuesta y otras partes interesadas.

Importancia

El Nivel 4: Búsqueda y Rescate Total es esencial para asegurar que todas las víctimas atrapadas o desaparecidas sean encontradas y rescatadas de manera segura y eficiente. Esta fase requiere una planificación y coordinación meticulosas, así como el uso de técnicas y equipos avanzados. La efectividad de las operaciones en esta fase puede significar la diferencia entre la vida y la muerte para las víctimas, por lo que es crucial que todos los recursos y esfuerzos se dirijan hacia este objetivo.

Transcripción.

El Nivel 4 en las operaciones de Urban Search and Rescue (USAR) se refiere a la fase de "Búsqueda y Rescate Total" (Full Search and Rescue). Esta fase es crucial y representa la culminación de los esfuerzos de evaluación y planificación previos. A continuación se describen las definiciones y el propósito del Nivel 4: Búsqueda y Rescate Total.

Definiciones

1. Búsqueda y Rescate Total:

- **Búsqueda:** La acción de localizar a las víctimas atrapadas o desaparecidas en el área afectada. Esto incluye el uso de técnicas avanzadas de búsqueda, como el uso de perros de rescate, equipos de detección de sonido y cámaras térmicas.
- **Rescate:** La acción de liberar y extraer a las víctimas de manera segura de las estructuras colapsadas o de otras situaciones peligrosas. Esto puede involucrar el uso de herramientas especializadas, como sierras, gatos hidráulicos y equipos de levantamiento.

Propósito

El propósito del Nivel 4: Búsqueda y Rescate Total es maximizar las posibilidades de encontrar y rescatar a todas las víctimas atrapadas o desaparecidas en el área afectada. Esta fase se caracteriza por una serie de actividades y objetivos específicos:

1. Localización de Víctimas:

- **Técnicas Avanzadas:** Utilizar técnicas avanzadas de búsqueda para localizar a las víctimas, incluyendo el uso de tecnología y equipos especializados.
- **Cobertura Completa:** Asegurarse de que todas las áreas del sitio de trabajo sean revisadas exhaustivamente para no dejar a ninguna víctima sin ser encontrada.

2. Extracción Segura:

- **Estabilización de Estructuras:** Estabilizar las estructuras dañadas para asegurar que los rescatistas puedan trabajar de manera segura.
- **Uso de Herramientas Especializadas:** Utilizar herramientas y equipos especializados para liberar y extraer a las víctimas de manera segura y eficiente.

3. Atención Médica Inmediata:

- **Evaluación Médica:** Evaluar el estado médico de las víctimas rescatadas y proporcionar atención médica inmediata según sea necesario.
- **Transporte a Instalaciones Médicas:** Coordinar el transporte de las víctimas a instalaciones médicas para recibir atención adicional.

4. Coordinación y Comunicación:

- **Coordinación de Equipos:** Coordinar los esfuerzos de los diferentes equipos de rescate para asegurar una respuesta integrada y eficiente.
- **Comunicación Continua:** Mantener una comunicación continua entre los equipos de rescate, los coordinadores de la respuesta y otras agencias involucradas.

5. Documentación y Registro:

- Registro de Víctimas: Documentar y registrar a todas las víctimas encontradas y rescatadas para mantener un registro preciso de la situación.
- Informes de Progreso: Proporcionar informes regulares sobre el progreso de las operaciones de búsqueda y rescate a los coordinadores de la respuesta y otras partes interesadas.

Importancia

El Nivel 4: Búsqueda y Rescate Total es esencial para asegurar que todas las víctimas atrapadas o desaparecidas sean encontradas y rescatadas de manera segura y eficiente. Esta fase requiere una planificación y coordinación meticulosas, así como el uso de técnicas y equipos avanzados. La efectividad de las operaciones en esta fase puede significar la diferencia entre la vida y la muerte para las víctimas, por lo que es crucial que todos los recursos y esfuerzos se dirijan hacia este objetivo.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.2159709618874773, recall=0.5891089108910891, fmeasure=0.3160690571049137)
ROUGE-2 score: Score(precision=0.05272727272727273, recall=0.14427860696517414, fmeasure=0.07723035952063914)
ROUGE-L score: Score(precision=0.11070780399274047, recall=0.30198019801980197, fmeasure=0.16201859229747678)
BERTScore - Precision: 0.6629543900489807, Recall: 0.6928144693374634, F1: 0.6775555610656738
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.216 indica que el 21.6% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.589 sugiere que el sistema recupera el 58.9% de las palabras relevantes.
- **F1-Score:** 0.316 refleja un **desequilibrio entre precisión y recall**, mostrando que el sistema es capaz de capturar una cantidad decente de contenido relevante, pero genera una cantidad significativa de contenido irrelevante.

2. ROUGE-2 Score:

- **Precisión:** 0.053 indica que solo el 5.3% de los bigrams generados son correctos.
- **Recall:** 0.144 muestra que el sistema recupera el 14.4% de los bigrams relevantes.
- **F1-Score:** 0.077 refleja un **rendimiento bajo** en la captura de secuencias de palabras significativas, con una **precisión limitada**.

3. ROUGE-L Score:

- **Precisión:** 0.111 sugiere que solo el 11.1% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.302 indica que el sistema recupera el 30.2% de la secuencia relevante más larga.
- **F1-Score:** 0.162 muestra un **rendimiento bajo en secuencias estructuradas**, con **baja precisión y recall limitado**.

4. BERTScore:

- **Precisión:** 0.663 indica que el 66.3% de las palabras generadas son semánticamente relevantes.

- **Recall:** 0.693 sugiere que el sistema captura el 69.3% del contenido semántico relevante.
- **F1-Score:** 0.678 refleja un **rendimiento semántico moderado**, con un **equilibrio aceptable** entre precisión y recall.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene un **recall moderado** pero una **precisión baja**, lo que sugiere que genera una cantidad significativa de contenido irrelevante, especialmente en la captura de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** muestra que el sistema es más sólido en términos de **semántica**, capturando el significado general de manera razonable, aunque aún con espacio para mejorar. Para mejorar el sistema, sería necesario aumentar la **precisión** y mejorar la **captura de secuencias coherentes** para generar respuestas más precisas y estructuradas.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

Este es el trabajo de SAR que debe identificar, ubicar y rescatar el pequeño número de sobrevivientes fuertemente atrapados o sepultados que los rescatistas locales, los primeros rescatistas, los recursos de LEMA o las operaciones ASR Nivel 3 no rescataron.

Los equipos penetran todos los vacíos penetrables, o la mayor parte de ellos.

Es probable que sean operaciones a un plazo más largo (más de un período operacional) que requieran un mayor rango de habilidades USAR, como: Todas las técnicas de búsqueda y equipos posibles de manera repetida a medida que se logre acceder.

Posiblemente, apuntalamiento extenso para hacer que las estructuras y las rutas de acceso sean más seguras.

Ruptura y separación fuerte y repetida de la totalidad de los elementos de la estructura.

Levantamiento y/o remoción de elementos grandes.

La eliminación de niveles de la estructura puede que sea necesaria en este nivel si es necesario obtener acceso para rescatar una víctima viva.

Trabajar en espacios confinados, algunas veces de manera profunda en las estructuras.

Varios equipos pueden estar involucrados en el mismo sitio de trabajo.

Se requiere mando y control total del sitio de trabajo.

Respuesta generada con RAG Básico:

Las definiciones y propósito del nivel 4: Búsqueda y rescate total son las siguientes:

- ****Definiciones y propósito:****

- Este nivel implica operaciones de búsqueda y rescate que son más extensas y complejas, generalmente ejecutadas después del nivel 3 o de manera conjunta con él.

- Incluye el uso de técnicas avanzadas de búsqueda y rescate, como la remoción de escombros, apuntalamiento, ruptura, separación, y penetración limitada en la estructura/escombros.

Anexo Emergencias y Ayuda Sanitaria.

TFM

Alumna: María Carolina Passarello.

- La búsqueda y/o rescate en este nivel puede completarse en un solo período operacional, a veces en unas pocas horas.
 - Los equipos no asumen operaciones a largo plazo para penetrar profundamente en la estructura, a menos que existan fuertes indicios de víctimas vivas.
 - Las víctimas profundamente sepultadas pueden no ser rescatadas en este nivel.
 - Los equipos deben identificar las estructuras o sitios de trabajo en los cuales una búsqueda de Nivel 4 se justifique.
 - Si se confirma una víctima viva atrapada profundamente, el equipo debe extender sus operaciones al Nivel 4 si las restricciones de la tarea lo permiten o si la coordinación del sector lo autoriza.
 - Los equipos deben asegurarse de cumplir con su trabajo de Nivel 3 en los sitios de trabajo restantes asignados a ellos. En caso de que no logren el rescate, deben solicitar recursos adicionales.
 - Si sitios de rescate adicionales son identificados en cualquier momento, entonces una nueva Identificación del Equipo debe ser creada.
 - SCC/UCC debe hacerles seguimiento a todas las tareas ASR3 y reasignar el sitio para ASR4, exceptuando los sitios reportados como no necesarios para ASR4.
- ****Ejecutado cuándo y por quién:****
- Este nivel es normalmente ejecutado después del Nivel 3, o de manera conjunta con él - Búsqueda y rescate rápido.
 - Si LEMA ya identificó sitios específicos, un equipo puede pasar directamente a las operaciones de Nivel 4 como su primera tarea.
 - Ejecutado por Equipos USAR Intermedios y Pesados.
- ****Herramientas INSARAG:****
- Formato Triage del Sitio de Trabajo.
 - Formato de Reporte del Sitio de Trabajo.
 - Formato de Liberación de Víctima.
 - Sistema de Marcaje del Sitio de Trabajo.
- ****Resultados:****
- Reportes diligenciados del Sitio de Trabajo.
 - Marcaje de los sitios de trabajo.
 - Formatos diligenciados de liberación de víctimas.

Métricas:

ROUGE-1 score: Score(precision=0.3368421052631579,
recall=0.6336633663366337, fmeasure=0.43986254295532645)
ROUGE-2 score: Score(precision=0.09498680738786279,
recall=0.1791044776119403, fmeasure=0.12413793103448274)
ROUGE-L score: Score(precision=0.13157894736842105,
recall=0.24752475247524752, fmeasure=0.1718213058419244)
BERTScore - Precision: 0.6726723313331604, Recall: 0.7124123573303223, F1:
0.6919723153114319

Análisis de las Métricas

- **ROUGE-1 Score:**

- **Precisión:** 0.337 indica que el 33.7% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.634 sugiere que el sistema recupera el 63.4% de las palabras relevantes.
- **F1-Score:** 0.440 refleja un **equilibrio aceptable** entre precisión y recall, aunque la precisión podría mejorarse para reducir la cantidad de contenido irrelevante.

- **ROUGE-2 Score:**

- **Precisión:** 0.095 indica que solo el 9.5% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.179 muestra que el sistema recupera el 17.9% de los bigrams relevantes.
- **F1-Score:** 0.124 refleja un **rendimiento bajo** en la captura de secuencias de palabras significativas, lo que sugiere que el sistema tiene dificultades para generar relaciones coherentes entre palabras.

- **ROUGE-L Score:**

- **Precisión:** 0.132 sugiere que solo el 13.2% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.248 indica que el sistema recupera el 24.8% de la secuencia relevante más larga.
- **F1-Score:** 0.172 refleja un **rendimiento bajo** en la generación de secuencias estructuradas.

- **BERTScore:**

- **Precisión:** 0.673 indica que el 67.3% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.712 sugiere que el sistema captura el 71.2% del contenido semántico relevante.
- **F1-Score:** 0.692 refleja un **rendimiento semántico moderado**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Respuesta generada con RAG con Haystack:

Definiciones y propósito del nivel 4: Búsqueda y rescate total

El nivel 4 de búsqueda y rescate total se define como el trabajo de búsqueda y rescate (SAR) que tiene como objetivo identificar, ubicar y rescatar a un pequeño número de sobrevivientes fuertemente atrapados o sepultados que no fueron rescatados por los rescatistas locales, los primeros rescatistas, los recursos de LEMA (Autoridad Local de Manejo de Emergencias) o las operaciones de nivel 3. Este nivel de operación es más intensivo y puede requerir un mayor rango de habilidades USAR (Urban Search and Rescue), tales como:

Anexo Emergencias y Ayuda Sanitaria.

TFM

Alumna: María Carolina Passarello.

- **Todas las técnicas de búsqueda y equipos posibles de manera repetida a medida que se logre acceder.**
- **Apuntalamiento extenso para hacer que las estructuras y las rutas de acceso sean más seguras.**
- **Ruptura y separación fuerte y repetida de la totalidad de los elementos de la estructura.**
- **Levantamiento y/o remoción de elementos grandes.**
- **Eliminación de niveles de la estructura puede que sea necesaria en este nivel si es necesario obtener acceso para rescatar una víctima viva.**
- **Trabajar en espacios confinados, algunas veces de manera profunda en las estructuras.**

Varios equipos pueden estar involucrados en el mismo sitio de trabajo. Este nivel de operación es probable que sea a un plazo más largo (más de un período operacional) y requiere un control y mando total del sitio de trabajo.

Ejecutado cuándo y por quién

1. **Este nivel es normalmente ejecutado después del Nivel 3, o de manera conjunta con él - Búsqueda y rescate rápido.**
2. **Si LEMA ya identificó sitios específicos, un equipo puede pasar directamente a las operaciones de Nivel 4 como su primera tarea.**
3. **Ejecutado por Equipos USAR Intermedios y Pesados.**

Herramientas INSARAG

1. **Formato Triage del Sitio de Trabajo.**
2. **Formato de Reporte del Sitio de Trabajo.**
3. **Formato de Liberación de Víctima.**
4. **Sistema de Marcaje del Sitio de Trabajo.**

Resultados

1. **Reportes diligenciados del Sitio de Trabajo.**
2. **Marcaje de los sitios de trabajo.**
3. **Formatos diligenciados de liberación de víctimas.**

Este nivel de operación es crucial para asegurar que todos los esfuerzos posibles se realicen para rescatar a las víctimas atrapadas, utilizando todas las técnicas y equipos disponibles.

Métricas:

ROUGE-1 score: Score(precision=0.4744897959183674,
recall=0.9207920792079208, fmeasure=0.62626262626263)
ROUGE-2 score: Score(precision=0.39641943734015345,
recall=0.7711442786069652, fmeasure=0.5236486486486487)
ROUGE-L score: Score(precision=0.40816326530612246,
recall=0.7920792079207921, fmeasure=0.5387205387205387)
BERTScore - Precision: 0.7161004543304443, Recall: 0.8216364979743958, F1:
0.7652469873428345

Análisis de las Métricas.

1. ROUGE-1 Score:

- **Precisión:** 0.474 indica que el 47.4% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.921 sugiere que el sistema recupera el 92.1% de las palabras relevantes, lo que indica un **recall muy alto**.
- **F1-Score:** 0.626 refleja un **buen equilibrio** entre precisión y recall, aunque la precisión podría mejorarse para reducir el contenido irrelevante.

2. ROUGE-2 Score:

- **Precisión:** 0.396 indica que el 39.6% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.771 muestra que el sistema recupera el 77.1% de los bigrams relevantes.
- **F1-Score:** 0.524 refleja un **buen rendimiento** en la captura de secuencias significativas, aunque hay margen para mejorar la precisión.

3. ROUGE-L Score:

- **Precisión:** 0.408 sugiere que el 40.8% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.792 muestra que el sistema recupera el 79.2% de la secuencia relevante más larga.
- **F1-Score:** 0.539 refleja un **buen rendimiento** en la generación de secuencias estructuradas, aunque la precisión podría mejorar.

4. BERTScore:

- **Precisión:** 0.716 indica que el 71.6% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.822 sugiere que el sistema captura el 82.2% del contenido semántico relevante.
- **F1-Score:** 0.765 refleja un **rendimiento semántico sólido**, con un buen equilibrio entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene un **recall muy alto**, pero la **precisión es moderada**, lo que sugiere que genera contenido relevante pero también una cantidad significativa de contenido irrelevante. El **rendimiento en secuencias** es bueno, aunque la precisión en la generación de secuencias (ROUGE-2 y ROUGE-L) podría mejorarse. El **BERTScore** muestra un **rendimiento semántico fuerte**, lo que indica que el sistema captura bien el significado general, pero podría beneficiarse de una mayor precisión literal para mejorar la relevancia general.

Respuesta generada con RAG con LlamaIndex

El nivel 4, Búsqueda y rescate total, implica operaciones en un sitio de trabajo para recuperar víctimas fallecidas. Es parte de la fase de rescate si es necesario por la autoridad coordinadora. Puede incluir la búsqueda en cada vacío posible, el uso de todas las habilidades USAR indicadas en el Nivel 4, la eliminación de niveles de la estructura para permitir acceso a todas las partes de la estructura o a pilas de escombros, y el

trabajo con maquinaria pesada como grúas y equipo de demolición para obtener acceso. El mando y control del sitio de trabajo es esencial.

Métricas

```
ROUGE-1 score: Score(precision=0.6886792452830188,
recall=0.3613861386138614, fmeasure=0.4740259740259741)
ROUGE-2 score: Score(precision=0.2761904761904762,
recall=0.14427860696517414, fmeasure=0.1895424836601307)
ROUGE-L score: Score(precision=0.3113207547169811,
recall=0.16336633663366337, fmeasure=0.2142857142857143)
BERTScore - Precision: 0.7439799904823303, Recall: 0.6952946782112122, F1:
0.718813955783844
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.689 indica que el 68.9% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.361 sugiere que el sistema recupera el 36.1% de las palabras relevantes.
- **F1-Score:** 0.474 refleja un **buen rendimiento en precisión**, pero un **recall limitado**, lo que significa que el sistema genera contenido relevante, pero no recupera la mayoría del contenido relevante disponible.

2. ROUGE-2 Score:

- **Precisión:** 0.276 indica que el 27.6% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.144 muestra que el sistema recupera solo el 14.4% de los bigrams relevantes.
- **F1-Score:** 0.190 refleja un **rendimiento bajo** en la captura de secuencias significativas, con una **precisión moderada** pero un **recall limitado**.

3. ROUGE-L Score:

- **Precisión:** 0.311 sugiere que el 31.1% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.163 indica que el sistema recupera el 16.3% de la secuencia relevante más larga.
- **F1-Score:** 0.214 refleja un **rendimiento bajo** en la generación de secuencias estructuradas, con un **buen balance en precisión** pero un **recall bajo**.

4. BERTScore:

- **Precisión:** 0.744 indica que el 74.4% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.695 sugiere que el sistema captura el 69.5% del contenido semántico relevante.
- **F1-Score:** 0.719 refleja un **rendimiento semántico sólido**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene una **alta precisión**, pero su **recall es limitado**, lo que sugiere que genera contenido relevante pero no logra recuperar todo el

contenido relevante disponible. El rendimiento en la generación de secuencias (ROUGE-2 y ROUGE-L) es moderado en precisión, pero bajo en recall. El **BERTScore** muestra que el sistema es **sólido en términos de semántica**, capturando bien el significado general. Para mejorar, sería necesario aumentar el **recall** y capturar más secuencias relevantes, manteniendo la precisión actual.

Pregunta 6: ¿Cuál es el objetivo de ASR 2?

Respuesta generada por LECHAT


Search...

Today

Como se compone el código d...

CP

Cuál es el objetivo de ASR 2?



El término "ASR 2" generalmente se refiere a "Assessment and Search Rescue 2" (Evaluación y Búsqueda y Rescate 2) en el contexto de las operaciones de Urban Search and Rescue (USAR). Este nivel se centra en la evaluación detallada del sitio y la implementación de estrategias de búsqueda y rescate más avanzadas y específicas. A continuación se describe el objetivo de ASR 2:

Objetivo de ASR 2
El objetivo principal de ASR 2 es realizar una evaluación más detallada y específica del sitio de trabajo para identificar y priorizar las áreas donde es más probable encontrar víctimas atrapadas o desaparecidas. Esta fase se enfoca en la implementación de técnicas de búsqueda y rescate más avanzadas y en la coordinación de recursos para maximizar la eficiencia y efectividad de las operaciones.

Componentes y Actividades de ASR 2

- Evaluación Detallada del Sitio:**
 - Mapeo y Documentación:** Realizar un mapeo detallado del área afectada, documentando la ubicación de estructuras colapsadas, peligros identificados y áreas de interés.
 - Evaluación Estructural:** Evaluar la estabilidad de las estructuras dañadas para determinar la seguridad de los equipos de rescate y las víctimas.
- Priorización de Áreas de Búsqueda:**
 - Identificación de Áreas de Alta Probabilidad:** Utilizar información recopilada durante la evaluación inicial para identificar áreas donde es más probable encontrar víctimas.
 - Asignación de Recursos:** Asignar recursos y equipos de rescate a las áreas prioritarias para maximizar las posibilidades de encontrar y rescatar víctimas.
- Implementación de Técnicas de Búsqueda Avanzadas:**
 - Uso de Tecnología:** Implementar tecnologías avanzadas como cámaras térmicas, equipos de detección de sonido y perros de rescate para localizar víctimas.
 - Búsqueda Sistemática:** Realizar búsquedas sistemáticas y metódicas en las áreas prioritarias para asegurar que todas las víctimas sean encontradas.
- Coordinación y Comunicación:**
 - Coordinación de Equipos:** Coordinar los esfuerzos de los diferentes equipos de rescate para asegurar una respuesta integrada y eficiente.

Como se compone el código d...

- **Comunicación Continua:** Mantener una comunicación continua entre los equipos de rescate, los coordinadores de la respuesta y otras agencias involucradas para asegurar que todos estén alineados con los objetivos y estrategias de búsqueda y rescate.

5. Planificación y Ejecución de Operaciones de Rescate:

- **Desarrollo de Planes de Rescate:** Desarrollar planes detallados para las operaciones de rescate, incluyendo la identificación de rutas de acceso y extracción.
- **Ejecución de Rescates:** Ejecutar operaciones de rescate utilizando técnicas y equipos especializados para liberar y extraer a las víctimas de manera segura.

Importancia de ASR 2

ASR 2 es una fase crítica en las operaciones de USAR, ya que permite a los equipos de rescate enfocar sus esfuerzos en las áreas donde es más probable encontrar víctimas. La evaluación detallada y la implementación de técnicas avanzadas de búsqueda y rescate aumentan significativamente las posibilidades de localizar y rescatar a las víctimas de manera eficiente y segura. La coordinación y comunicación efectivas durante esta fase son esenciales para asegurar que todos los recursos se utilicen de manera óptima y que las operaciones de rescate se realicen de manera ordenada y organizada.

Transcripción.

El término "ASR 2" generalmente se refiere a "Assessment and Search Rescue 2" (Evaluación y Búsqueda y Rescate 2) en el contexto de las operaciones de Urban Search and Rescue (USAR). Este nivel se centra en la evaluación detallada del sitio y la implementación de estrategias de búsqueda y rescate más avanzadas y específicas. A continuación se describe el objetivo de ASR 2:

Objetivo de ASR 2

El objetivo principal de ASR 2 es realizar una evaluación más detallada y específica del sitio de trabajo para identificar y priorizar las áreas donde es más probable encontrar víctimas atrapadas o desaparecidas. Esta fase se enfoca en la implementación de técnicas de búsqueda y rescate más avanzadas y en la coordinación de recursos para maximizar la eficiencia y efectividad de las operaciones.

Componentes y Actividades de ASR 2

1. Evaluación Detallada del Sitio:

- **Mapeo y Documentación:** Realizar un mapeo detallado del área afectada, documentando la ubicación de estructuras colapsadas, peligros identificados y áreas de interés.
- **Evaluación Estructural:** Evaluar la estabilidad de las estructuras dañadas para determinar la seguridad de los equipos de rescate y las víctimas.

2. Priorización de Áreas de Búsqueda:

- **Identificación de Áreas de Alta Probabilidad:** Utilizar información recopilada durante la evaluación inicial para identificar áreas donde es más probable encontrar víctimas.
- **Asignación de Recursos:** Asignar recursos y equipos de rescate a las áreas prioritarias para maximizar las posibilidades de encontrar y rescatar víctimas.

3. Implementación de Técnicas de Búsqueda Avanzadas:

- **Uso de Tecnología:** Implementar tecnologías avanzadas como cámaras térmicas, equipos de detección de sonido y perros de rescate para localizar víctimas.

- **Búsqueda Sistemática:** Realizar búsquedas sistemáticas y metódicas en las áreas prioritarias para asegurar que todas las víctimas sean encontradas.

4. Coordinación y Comunicación:

- **Coordinación de Equipos:** Coordinar los esfuerzos de los diferentes equipos de rescate para asegurar una respuesta integrada y eficiente.
- **Comunicación Continua:** Mantener una comunicación continua entre los equipos de rescate, los coordinadores de la respuesta y otras agencias involucradas para asegurar que todos estén alineados con los objetivos y estrategias de búsqueda y rescate.

5. Planificación y Ejecución de Operaciones de Rescate:

- **Desarrollo de Planes de Rescate:** Desarrollar planes detallados para las operaciones de rescate, incluyendo la identificación de rutas de acceso y extracción.
- **Ejecución de Rescates:** Ejecutar operaciones de rescate utilizando técnicas y equipos especializados para liberar y extraer a las víctimas de manera segura.

Importancia de ASR 2

ASR 2 es una fase crítica en las operaciones de USAR, ya que permite a los equipos de rescate enfocar sus esfuerzos en las áreas donde es más probable encontrar víctimas. La evaluación detallada y la implementación de técnicas avanzadas de búsqueda y rescate aumentan significativamente las posibilidades de localizar y rescatar a las víctimas de manera eficiente y segura. La coordinación y comunicación efectivas durante esta fase son esenciales para asegurar que todos los recursos se utilicen de manera óptima y que las operaciones de rescate se realicen de manera ordenada y organizada.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.08348134991119005,
recall=0.7121212121212122, fmeasure=0.14944356120826707)
ROUGE-2 score: Score(precision=0.030249110320284697,
recall=0.26153846153846155, fmeasure=0.05422647527910686)
ROUGE-L score: Score(precision=0.0674955595026643, recall=0.5757575757575758,
fmeasure=0.12082670906200318)
BERTScore - Precision: 0.6526629328727722, Recall: 0.7314852476119995, F1:
0.6898297667503357
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.083 indica que solo el 8.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.712 sugiere que el sistema recupera el 71.2% de las palabras relevantes, lo que refleja un **recall alto**, pero una **precisión muy baja**.
- **F1-Score:** 0.149 refleja un **desequilibrio significativo** entre precisión y recall, lo que sugiere que el sistema genera mucho contenido irrelevante.

2. ROUGE-2 Score:

- **Precisión:** 0.030 indica que solo el 3.0% de los bigrams generados son correctos en comparación con el texto de referencia.
 - **Recall:** 0.262 muestra que el sistema recupera el 26.2% de los bigrams relevantes.
 - **F1-Score:** 0.054 refleja un **rendimiento bajo** en la captura de secuencias de palabras significativas, con una **precisión extremadamente limitada**.
3. **ROUGE-L Score:**
- **Precisión:** 0.067 sugiere que solo el 6.7% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.576 muestra que el sistema recupera el 57.6% de la secuencia relevante más larga.
 - **F1-Score:** 0.121 refleja un **rendimiento bajo** en la generación de secuencias estructuradas, con un **bajo nivel de precisión**.
4. **BERTScore:**
- **Precisión:** 0.653 indica que el 65.3% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.731 sugiere que el sistema captura el 73.1% del contenido semántico relevante.
 - **F1-Score:** 0.690 refleja un **rendimiento semántico razonable**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran un **recall alto**, pero una **precisión extremadamente baja**, lo que significa que el sistema genera una cantidad considerable de contenido irrelevante, especialmente en la captura de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** indica un **rendimiento semántico moderado**, lo que sugiere que el sistema captura bien el significado general, aunque necesita mejorar en **precisión literal** y en la **estructura de las secuencias** para generar respuestas más concisas y coherentes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

El objetivo de ASR 2 Evaluación del Sitio de Trabajo es evaluar las estructuras colapsadas e identificar sitios viables para el rescate de víctimas. UCC usa esta información para hacer listados de sitios, darles prioridad y decidir qué equipos asignarles a cuáles sitios. Una de las consideraciones para la priorización de los sitios de trabajo es la categoría de triage.

Respuesta generada con RAG Básico:

El objetivo de ASR 2 es la Evaluación del Triage del sitio de trabajo.

Métricas:

ROUGE-1 score: Score(precision=0.9333333333333333,
recall=0.21212121212121213, fmeasure=0.345679012345679)

```
ROUGE-2 score: Score(precision=0.7142857142857143,  
recall=0.15384615384615385, fmeasure=0.25316455696202533)  
ROUGE-L score: Score(precision=0.7333333333333333,  
recall=0.16666666666666666, fmeasure=0.2716049382716049)  
BERTScore - Precision: 0.8494949340820312, Recall: 0.6948264241218567, F1:  
0.764415442943573
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.933 indica que el 93.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.212 sugiere que el sistema recupera solo el 21.2% de las palabras relevantes.
- **F1-Score:** 0.346 refleja un **desequilibrio** significativo entre precisión y recall, lo que indica que el sistema genera contenido muy relevante pero recupera solo una pequeña parte de la información disponible.

2. ROUGE-2 Score:

- **Precisión:** 0.714 indica que el 71.4% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.154 muestra que el sistema recupera solo el 15.4% de los bigrams relevantes.
- **F1-Score:** 0.253 refleja un **rendimiento moderado** en la captura de secuencias significativas, con una **alta precisión** pero un **recall bajo**.

3. ROUGE-L Score:

- **Precisión:** 0.733 sugiere que el 73.3% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.167 indica que el sistema recupera solo el 16.7% de la secuencia relevante más larga.
- **F1-Score:** 0.272 refleja un **bajo rendimiento** en la captura de secuencias estructuradas, con una **alta precisión** pero un **recall limitado**.

4. BERTScore:

- **Precisión:** 0.849 indica que el 84.9% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.695 sugiere que el sistema captura el 69.5% del contenido semántico relevante.
- **F1-Score:** 0.764 refleja un **buen rendimiento semántico**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene una **precisión muy alta**, pero su **recall es limitado**, lo que significa que genera contenido muy relevante, pero no logra recuperar toda la información disponible. El **BERTScore** sugiere que el sistema es **sólido en términos de semántica**, capturando bien el significado general.

Respuesta generada con RAG con Haystack:

El objetivo de ASR 2 es evaluar las estructuras colapsadas e identificar sitios viables para el rescate de víctimas. UCC utiliza esta información para hacer listados de sitios, darles prioridad y decidir qué equipos asignarles a cuáles sitios. Una de las consideraciones para la priorización de los sitios de trabajo es la categoría de triage.

Métricas:

ROUGE-1 score: Score(precision=0.9833333333333333, recall=0.8939393939393939, fmeasure=0.9365079365079364)
ROUGE-2 score: Score(precision=0.9491525423728814, recall=0.8615384615384616, fmeasure=0.903225806451613)
ROUGE-L score: Score(precision=0.9833333333333333, recall=0.8939393939393939, fmeasure=0.9365079365079364)
BERTScore - Precision: 0.9817554354667664, Recall: 0.9565548300743103, F1: 0.9689912796020508

Análisis de las Métricas.

1. ROUGE-1 Score:

- **Precisión:** 0.983 indica que el 98.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.894 sugiere que el sistema recupera el 89.4% de las palabras relevantes, lo que refleja un **muy buen recall**.
- **F1-Score:** 0.937 muestra un **equilibrio excelente** entre precisión y recall, lo que indica que el sistema genera contenido muy relevante y recupera la mayor parte de la información disponible.

2. ROUGE-2 Score:

- **Precisión:** 0.949 indica que el 94.9% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.862 sugiere que el sistema recupera el 86.2% de los bigrams relevantes.
- **F1-Score:** 0.903 refleja un **rendimiento sobresaliente** en la captura de secuencias significativas, con una **alta precisión y buen recall**.

3. ROUGE-L Score:

- **Precisión:** 0.983 sugiere que el 98.3% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.894 indica que el sistema recupera el 89.4% de la secuencia relevante más larga.
- **F1-Score:** 0.937 muestra un **rendimiento excepcional** en la captura de secuencias estructuradas, con un **equilibrio excelente** entre precisión y recall.

4. BERTScore:

- **Precisión:** 0.982 indica que el 98.2% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.957 sugiere que el sistema captura el 95.7% del contenido semántico relevante.
- **F1-Score:** 0.969 refleja un **rendimiento semántico casi perfecto**, con un **equilibrio sólido** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican un **rendimiento sobresaliente**, tanto en precisión como en recall. El sistema es **extremadamente eficaz** en la generación de contenido relevante y en la captura de secuencias significativas y semánticamente coherentes. La **precisión y el recall son ambos altos**, lo que sugiere que el sistema es muy eficiente en la generación de respuestas completas y precisas. En resumen, el sistema tiene un **rendimiento excepcional** tanto en términos de literalidad como de significado.

Respuesta generada con RAG con LLaIndex

El objetivo de ASR 2 es identificar y ubicar a las víctimas atrapadas en estructuras colapsadas o en escombros, y proporcionarles asistencia inmediata para su rescate. Este nivel de búsqueda y rescate se enfoca en la rápida localización y extracción de sobrevivientes, utilizando técnicas y equipos básicos de rescate.

Métricas

```
ROUGE-1 score: Score(precision=0.4107142857142857,
recall=0.3484848484848485, fmeasure=0.3770491803278689)
ROUGE-2 score: Score(precision=0.127272727272726,
recall=0.1076923076923077, fmeasure=0.11666666666666667)
ROUGE-L score: Score(precision=0.30357142857142855,
recall=0.25757575757575757, fmeasure=0.2786885245901639)
BERTScore - Precision: 0.7663599848747253, Recall: 0.7351076602935791, F1:
0.7504085898399353
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.411 indica que el 41.1% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.348 sugiere que el sistema recupera el 34.8% de las palabras relevantes.
- **F1-Score:** 0.377 refleja un **equilibrio moderado** entre precisión y recall, aunque ambos valores podrían mejorarse para aumentar la relevancia del contenido generado.

2. ROUGE-2 Score:

- **Precisión:** 0.127 indica que solo el 12.7% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.108 sugiere que el sistema recupera solo el 10.8% de los bigrams relevantes.
- **F1-Score:** 0.117 refleja un **rendimiento bajo** en la captura de secuencias significativas, lo que sugiere que el sistema tiene dificultades para generar relaciones coherentes entre palabras.

3. ROUGE-L Score:

- **Precisión:** 0.304 sugiere que el 30.4% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.

- **Recall:** 0.258 indica que el sistema recupera el 25.8% de la secuencia relevante más larga.
 - **F1-Score:** 0.279 refleja un **rendimiento bajo en secuencias estructuradas**, con una precisión y recall limitados.
4. **BERTScore:**
- **Precisión:** 0.766 indica que el 76.6% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.735 sugiere que el sistema captura el 73.5% del contenido semántico relevante.
 - **F1-Score:** 0.750 refleja un **buen rendimiento semántico**, con un **equilibrio aceptable** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **rendimiento moderado** en la generación de palabras relevantes, pero **bajo en la captura de secuencias significativas** (ROUGE-2 y ROUGE-L), lo que sugiere que el sistema tiene dificultades para generar secuencias coherentes. El **BERTScore** muestra un **rendimiento semántico sólido**, lo que sugiere que, aunque el sistema capta bien el significado general, necesita mejorar en **precisión literal** y en la **estructura de secuencias** para generar respuestas más precisas y coherentes.

Pregunta 7: ¿Cuál es el objetivo del proceso de triage?

Respuesta generada por LECHAT

Search...

Today

Como se compone el código d...

Search...

Today

Como se compone el código d...

Today

Como se compone el código d...

CP

cuál es el objetivo del proceso de triage?

M

El objetivo del proceso de triage en el contexto de las operaciones de Urban Search and Rescue (USAR) y en situaciones de emergencia en general es priorizar la atención médica y los recursos de rescate basándose en la gravedad de las lesiones y la urgencia de la atención necesaria. Este proceso es crucial para maximizar la eficiencia y efectividad de la respuesta en situaciones donde los recursos pueden ser limitados y el tiempo es esencial. A continuación se describen los objetivos específicos del proceso de triage:

Objetivos del Proceso de Triage

1. Priorización de la Atención Médica:

Identificación de Víctimas Críticas: Identificar rápidamente a las víctimas que requieren atención médica inmediata debido a lesiones graves o condiciones críticas.

Clasificación de Víctimas: Clasificar a las víctimas en categorías de prioridad (por ejemplo, rojo para víctimas críticas, amarillo para víctimas con lesiones graves pero no inmediatamente mortales, verde para víctimas con lesiones menores, y negro para víctimas fallecidas) para guiar la atención médica.

2. Optimización de Recursos:

Asignación Eficiente de Recursos: Asegurar que los recursos médicos y de rescate se asignen de manera eficiente y efectiva, priorizando a las víctimas que más lo necesitan.

Gestión de Recursos Limitados: Maximizar el uso de recursos limitados, como personal médico, equipos de rescate y suministros, en situaciones de emergencia.

3. Reducción de la Morbilidad y Mortalidad:

Atención Rápida a Víctimas Críticas: Proporcionar atención médica rápida y adecuada a las víctimas críticas para aumentar sus posibilidades de supervivencia.

Prevención de Complicaciones: Identificar y tratar lesiones que podrían empeorar si no se atienden de manera oportuna, reduciendo así la morbilidad y mortalidad.

4. Coordinación y Comunicación:

Información Precisa: Proporcionar información precisa y actualizada sobre el estado de las víctimas a los equipos de rescate y coordinadores de la respuesta.

Coordinación de Atención: Facilitar la coordinación entre los equipos médicos, los equipos de rescate y otras agencias involucradas para asegurar una respuesta integrada y eficiente.

5. Documentación y Registro:

Registro de Víctimas: Documentar y registrar a todas las víctimas evaluadas y clasificadas para mantener un registro preciso de la situación.

Informes de Progreso: Proporcionar informes regulares sobre el progreso del triage y la atención médica a los coordinadores de la respuesta y otras partes interesadas.

Importancia del Proceso de Triage

El proceso de triage es fundamental en situaciones de emergencia y desastres, ya que permite a los equipos de rescate y médicos tomar decisiones informadas y coordinar sus esfuerzos de manera efectiva. En situaciones donde los recursos son limitados y el tiempo es crítico, el triage asegura que las víctimas más necesitadas reciban la atención adecuada de manera oportuna, maximizando así las posibilidades de salvar vidas y reducir la morbilidad.

En resumen, el objetivo del proceso de triage es priorizar la atención médica y los recursos de rescate basándose en la gravedad de las lesiones y la urgencia de la atención necesaria, optimizando así la eficiencia y efectividad de la respuesta en situaciones de emergencia.

Transcripción.

El objetivo del proceso de triage en el contexto de las operaciones de Urban Search and Rescue (USAR) y en situaciones de emergencia en general es priorizar la atención médica y los recursos de rescate basándose en la gravedad de las lesiones y la urgencia de la atención necesaria. Este proceso es crucial para maximizar la eficiencia y efectividad de la respuesta en situaciones donde los recursos pueden ser limitados y el tiempo es esencial. A continuación se describen los objetivos específicos del proceso de triage:

Objetivos del Proceso de Triage

1. Priorización de la Atención Médica:

- **Identificación de Víctimas Críticas:** Identificar rápidamente a las víctimas que requieren atención médica inmediata debido a lesiones graves o condiciones críticas.

- **Clasificación de Víctimas:** Clasificar a las víctimas en categorías de prioridad (por ejemplo, rojo para víctimas críticas, amarillo para víctimas con lesiones graves pero no inmediatamente mortales, verde para víctimas con lesiones menores, y negro para víctimas fallecidas) para guiar la atención médica.

2. Optimización de Recursos:

- **Asignación Eficiente de Recursos:** Asegurar que los recursos médicos y de rescate se asignen de manera eficiente y efectiva, priorizando a las víctimas que más lo necesitan.

- **Gestión de Recursos Limitados:** Maximizar el uso de recursos limitados, como personal médico, equipos de rescate y suministros, en situaciones de emergencia.

3. Reducción de la Morbilidad y Mortalidad:

- **Atención Rápida a Víctimas Críticas:** Proporcionar atención médica rápida y adecuada a las víctimas críticas para aumentar sus posibilidades de supervivencia.

- **Prevención de Complicaciones:** Identificar y tratar lesiones que podrían empeorar si no se atienden de manera oportuna, reduciendo así la morbilidad y mortalidad.

4. Coordinación y Comunicación:

- **Información Precisa:** Proporcionar información precisa y actualizada sobre el estado de las víctimas a los equipos de rescate y coordinadores de la respuesta.

- **Coordinación de Atención:** Facilitar la coordinación entre los equipos médicos, los equipos de rescate y otras agencias involucradas para asegurar una respuesta integrada y eficiente.

5. Documentación y Registro:

- **Registro de Víctimas:** Documentar y registrar a todas las víctimas evaluadas y clasificadas para mantener un registro preciso de la situación.

- **Informes de Progreso:** Proporcionar informes regulares sobre el progreso del triage y la atención médica a los coordinadores de la respuesta y otras partes interesadas.

Importancia del Proceso de Triage

El proceso de triage es fundamental en situaciones de emergencia y desastres, ya que permite a los equipos de rescate y médicos tomar decisiones informadas y coordinar sus esfuerzos de manera efectiva. En situaciones donde los recursos son limitados y el tiempo es crítico, el triage asegura que las víctimas más necesitadas reciban la atención adecuada de manera oportuna, maximizando así las posibilidades de salvar vidas y reducir la morbilidad.

En resumen, el objetivo del proceso de triage es priorizar la atención médica y los recursos de rescate basándose en la gravedad de las lesiones y la urgencia de la atención necesaria, optimizando así la eficiencia y efectividad de la respuesta en situaciones de emergencia.

Métricas sin RAG.

ROUGE-1 score: Score(precision=0.05197132616487455, recall=0.725, fmeasure=0.09698996655518394)

ROUGE-2 score: Score(precision=0.02333931777378815, recall=0.3333333333333333, fmeasure=0.0436241610738255)

ROUGE-L score: Score(precision=0.04838709677419355, recall=0.675, fmeasure=0.0903010033444816)

BERTScore - Precision: 0.6007959246635437, Recall: 0.7585302591323853, F1: 0.6705114841461182

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.052 indica que solo el 5.2% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.725 sugiere que el sistema recupera el 72.5% de las palabras relevantes, lo que refleja un **recall alto**, pero con una **precisión muy baja**.
- **F1-Score:** 0.097 muestra un **desequilibrio significativo** entre precisión y recall, lo que indica que el sistema genera mucho contenido irrelevante.

2. ROUGE-2 Score:

- **Precisión:** 0.023 indica que solo el 2.3% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.333 sugiere que el sistema recupera el 33.3% de los bigrams relevantes.
- **F1-Score:** 0.044 refleja un **rendimiento muy bajo** en la captura de secuencias significativas, con una **precisión extremadamente baja**.

3. ROUGE-L Score:

- **Precisión:** 0.048 sugiere que solo el 4.8% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.675 muestra que el sistema recupera el 67.5% de la secuencia relevante más larga.
- **F1-Score:** 0.090 refleja un **rendimiento bajo** en la captura de secuencias estructuradas, con un **recall alto**, pero una **precisión muy limitada**.

4. BERTScore:

- **Precisión:** 0.601 indica que el 60.1% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.759 sugiere que el sistema captura el 75.9% del contenido semántico relevante.
- **F1-Score:** 0.671 refleja un **rendimiento semántico aceptable**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran un **recall alto**, pero una **precisión extremadamente baja**, lo que sugiere que el sistema genera mucho contenido irrelevante, especialmente en la captura de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** sugiere un **rendimiento semántico moderado**, lo que indica que el sistema capta razonablemente bien el significado general, aunque necesita mejorar considerablemente en **precisión literal** y en la **estructura de las secuencias** para generar respuestas más coherentes y relevantes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

El objetivo del proceso de triage es evaluar los factores de triage para comparar las estructuras colapsadas y decidir el orden de las prioridades. La clave para el triage es consistencia en la comparación de los factores de triage.

Respuesta generada con RAG Básico:

El objetivo del proceso de triage es evaluar los factores de triage para comparar las estructuras colapsadas y decidir el orden de las prioridades. La clave para el triage es la consistencia en la comparación de los factores de triage.

Métricas:

ROUGE-1 score: Score(precision=0.975609756097561, recall=1.0, fmeasure=0.9876543209876543)
ROUGE-2 score: Score(precision=0.95, recall=0.9743589743589743, fmeasure=0.9620253164556962)
ROUGE-L score: Score(precision=0.975609756097561, recall=1.0, fmeasure=0.9876543209876543)
BERTScore - Precision: 0.9943650960922241, Recall: 0.997122585773468, F1: 0.9957419037818909

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.976 indica que el 97.6% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 1.0 sugiere que el sistema recupera el 100% de las palabras relevantes, lo que refleja un **recall perfecto**.
- **F1-Score:** 0.988 muestra un **rendimiento sobresaliente**, con un equilibrio casi perfecto entre precisión y recall.

2. ROUGE-2 Score:

- **Precisión:** 0.950 indica que el 95.0% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.974 muestra que el sistema recupera el 97.4% de los bigrams relevantes.
- **F1-Score:** 0.962 refleja un **rendimiento excepcional** en la captura de secuencias significativas, con **alta precisión y recall**.

3. ROUGE-L Score:

- **Precisión:** 0.976 indica que el 97.6% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 1.0 refleja que el sistema recupera la secuencia relevante más larga en su totalidad.
- **F1-Score:** 0.988 muestra un **rendimiento sobresaliente** en la captura de secuencias estructuradas.

4. BERTScore:

- **Precisión:** 0.994 indica que el 99.4% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.997 sugiere que el sistema captura el 99.7% del contenido semántico relevante.
- **F1-Score:** 0.996 refleja un **rendimiento semántico casi perfecto**, con un equilibrio excepcional entre precisión y recall.

Conclusión: Las métricas **ROUGE** y **BERTScore** indican un **rendimiento excepcional** del sistema en términos de precisión y recall. El sistema es **extremadamente eficaz en la generación de contenido relevante**, tanto en palabras como en secuencias. La **precisión y el recall** son casi perfectos, lo que sugiere que el sistema genera respuestas **muy precisas y completas**, con una correspondencia semántica excelente.

Respuesta generada con RAG con Haystack:

El objetivo del proceso de triage es evaluar los factores de triage para comparar las estructuras colapsadas y decidir el orden de las prioridades. La clave para el triage es la consistencia en la comparación de los factores de triage.

Métricas:

ROUGE-1 score: Score(precision=0.975609756097561, recall=1.0, fmeasure=0.9876543209876543)
ROUGE-2 score: Score(precision=0.95, recall=0.9743589743589743, fmeasure=0.9620253164556962)
ROUGE-L score: Score(precision=0.975609756097561, recall=1.0, fmeasure=0.9876543209876543)
BERTScore - Precision: 0.9943650960922241, Recall: 0.997122585773468, F1: 0.9957419037818909

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 97.6%, lo que significa que casi todas las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 100%, lo que indica que el sistema ha recuperado todas las palabras relevantes.
- **F1-Score:** 98.8%, mostrando un equilibrio excelente entre precisión y recall, con un rendimiento sobresaliente.

2. ROUGE-2 Score:

- **Precisión:** 95.0%, lo que refleja que la gran mayoría de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 97.4%, lo que indica que el sistema recupera casi todos los bigrams relevantes.
- **F1-Score:** 96.2%, lo que sugiere un rendimiento casi perfecto en la captura de secuencias de palabras.

3. ROUGE-L Score:

- **Precisión:** 97.6% y **recall** del 100%, lo que muestra que el sistema ha capturado las secuencias más largas de coincidencias de manera casi perfecta.
 - **F1-Score:** 98.8%, reflejando un rendimiento excelente en términos de estructura de secuencias.
4. **BERTScore:**
- **Precisión:** 99.4%, lo que indica una altísima relevancia semántica en las palabras generadas.
 - **Recall:** 99.7%, lo que muestra que casi todo el contenido semántico relevante fue recuperado.
 - **F1-Score:** 99.6%, lo que sugiere una correspondencia semántica casi perfecta.

Conclusión: El sistema muestra un **rendimiento sobresaliente en todas las métricas**, con una **altísima precisión y recall**. Esto indica que genera respuestas **extremadamente precisas, completas y coherentes**, capturando no solo las palabras literales, sino también el significado semántico de manera casi perfecta. El sistema es muy eficaz para generar contenido relevante, tanto en términos literales como semánticos.

Respuesta generada con RAG con LLaIndex

El objetivo del proceso de triage es evaluar los factores de triage para comparar las estructuras colapsadas y decidir el orden de las prioridades. La clave para el triage es la consistencia en la comparación de los factores de triage.

Métricas

```
ROUGE-1 score: Score(precision=0.975609756097561, recall=1.0,
fmeasure=0.9876543209876543)
ROUGE-2 score: Score(precision=0.95, recall=0.9743589743589743,
fmeasure=0.9620253164556962)
ROUGE-L score: Score(precision=0.975609756097561, recall=1.0,
fmeasure=0.9876543209876543)
BERTScore - Precision: 0.9943650960922241, Recall: 0.997122585773468, F1:
0.9957419037818909
```

Análisis de las Métricas

1. **ROUGE-1 Score:**
 - **Precisión:** 97.6%, lo que significa que casi todas las palabras generadas son relevantes.
 - **Recall:** 100%, lo que sugiere que el sistema recupera todas las palabras relevantes.
 - **F1-Score:** 98.8%, lo que refleja un equilibrio excelente entre precisión y recall, con un rendimiento casi perfecto.
2. **ROUGE-2 Score:**

- **Precisión:** 95.0%, lo que indica que la gran mayoría de los bigrams generados coinciden con el texto de referencia.
 - **Recall:** 97.4%, lo que significa que el sistema recupera la mayoría de los bigrams relevantes.
 - **F1-Score:** 96.2%, lo que demuestra un rendimiento excelente en la generación de secuencias significativas.
3. **ROUGE-L Score:**
- **Precisión:** 97.6%, lo que significa que el sistema es preciso en la captura de la secuencia más larga de coincidencias.
 - **Recall:** 100%, lo que refleja que el sistema recupera todas las secuencias relevantes.
 - **F1-Score:** 98.8%, lo que indica un rendimiento sobresaliente en la estructura de secuencias.
4. **BERTScore:**
- **Precisión:** 99.4%, lo que indica que el sistema genera contenido semánticamente relevante casi en su totalidad.
 - **Recall:** 99.7%, lo que significa que prácticamente todo el contenido semántico relevante fue recuperado.
 - **F1-Score:** 99.6%, lo que sugiere una correspondencia semántica casi perfecta.

Conclusión: El sistema tiene un **rendimiento casi perfecto** tanto en términos de precisión como de recall. Captura las palabras, secuencias y significados de manera extremadamente precisa y completa. Este equilibrio perfecto entre precisión y recall indica que el sistema genera contenido **altamente relevante, coherente y semánticamente preciso**.

Pregunta 8: En la categoría de Triage A, ¿Cuál es la duración esperada de la operación?

Respuesta generada por LECHAT

The screenshot shows the LECHAT interface. On the left, there is a sidebar with a search bar containing 'Search...', a 'Today' section, and a prompt 'Como se compone el código d...'. The main area displays the results of a search for 'Triage A'. It includes a brief description of Triage A (also known as Triage 1 or Triage Rojo) as victims with critical injuries needing immediate medical attention. Below this, there is a section titled 'Duración Esperada de la Operación para Triage A' which states that the duration is extremely short due to the need for immediate medical attention to increase survival chances. A sub-section 'Factores que Influyen en la Duración de la Operación:' lists '1. Gravedad de las Lesiones' as a key factor, explaining that severe injuries like hemorrhages or respiratory difficulties require immediate medical intervention.

Search...

Today

Como se compone el código d...

Today

Como se compone el código d...

2. Disponibilidad de Recursos Médicos:

- La disponibilidad de personal médico, equipos de rescate y suministros es crucial para proporcionar atención inmediata a las víctimas en Triage A. En situaciones de recursos limitados, la atención puede tardar más tiempo, lo cual es crítico para estas víctimas.

3. Condiciones del Incidente:

- Las condiciones específicas del incidente, como la estabilidad de las estructuras, la accesibilidad al sitio y la presencia de peligros adicionales, pueden influir en la rapidez con la que se puede atender a las víctimas en Triage A.

4. Coordinación y Comunicación:

- La eficiencia de la coordinación y comunicación entre los equipos de rescate, los equipos médicos y los coordinadores de la respuesta es esencial para asegurar que las víctimas en Triage A reciban atención inmediata.

Estimación General de la Duración

En términos generales, se espera que las víctimas en Triage A reciban atención médica inmediata, idealmente dentro de los primeros minutos después de ser clasificadas. Este marco de tiempo es crucial para estabilizar a las víctimas y prevenir complicaciones fatales. La rapidez en la intervención es esencial para aumentar las posibilidades de supervivencia de estas víctimas.

Importancia de la Atención Inmediata

La atención inmediata a las víctimas en Triage A es vital para prevenir complicaciones y asegurar su estabilización. La demora en la atención puede resultar en la pérdida de vidas, ya que estas víctimas están en una condición crítica que pone en riesgo su vida. Por lo tanto, es esencial que los equipos de rescate y médicos trabajen de manera coordinada y eficiente para proporcionar atención a estas víctimas dentro de un marco de tiempo extremadamente corto.

En resumen, la duración esperada de la operación para las víctimas en Triage A es extremadamente corta, generalmente dentro de los primeros minutos después de ser clasificadas. La rapidez en la intervención es crucial para estabilizar a las víctimas y prevenir complicaciones fatales, aumentando así sus posibilidades de supervivencia.

Transcripción.

En el contexto de las operaciones de Urban Search and Rescue (USAR) y en situaciones de emergencia, la categoría de Triage A (también conocida como Triage 1 o Triage Rojo) se refiere a las víctimas que tienen lesiones críticas y necesitan atención médica inmediata para sobrevivir. Estas víctimas están en una condición que pone en riesgo su vida y requieren intervención médica urgente.

Duración Esperada de la Operación para Triage A

La duración esperada de la operación para las víctimas clasificadas en Triage A es extremadamente corta, ya que estas víctimas necesitan atención médica inmediata para aumentar sus posibilidades de supervivencia. La rapidez en la intervención es crucial para estabilizar su condición y prevenir complicaciones fatales.

Factores que Influyen en la Duración de la Operación:

1. Gravedad de las Lesiones:

- Las lesiones de las víctimas en Triage A pueden incluir hemorragias graves, traumas severos, dificultades respiratorias, etc. La gravedad de estas lesiones requiere una intervención médica inmediata para estabilizar a la víctima.

2. Disponibilidad de Recursos Médicos:

- La disponibilidad de personal médico, equipos de rescate y suministros es crucial para proporcionar atención inmediata a las víctimas en Triage A. En situaciones de recursos limitados, la atención puede tardar más tiempo, lo cual es crítico para estas víctimas.

3. Condiciones del Incidente:

- Las condiciones específicas del incidente, como la estabilidad de las estructuras, la accesibilidad al sitio y la presencia de peligros adicionales, pueden influir en la rapidez con la que se puede atender a las víctimas en Triage A.

4. Coordinación y Comunicación:

- La eficiencia de la coordinación y comunicación entre los equipos de rescate, los equipos médicos y los coordinadores de la respuesta es esencial para asegurar que las víctimas en Triage A reciban atención inmediata.

Estimación General de la Duración

En términos generales, se espera que las víctimas en Triage A reciban atención médica inmediata, idealmente dentro de los primeros minutos después de ser clasificadas. Este marco de tiempo es crucial para estabilizar a las víctimas y prevenir complicaciones fatales. La rapidez en la intervención es esencial para aumentar las posibilidades de supervivencia de estas víctimas.

Importancia de la Atención Inmediata

La atención inmediata a las víctimas en Triage A es vital para prevenir complicaciones y asegurar su estabilización. La demora en la atención puede resultar en la pérdida de vidas, ya que estas víctimas están en una condición crítica que pone en riesgo su vida. Por lo tanto, es esencial que los equipos de rescate y médicos trabajen de manera coordinada y eficiente para proporcionar atención a estas víctimas dentro de un marco de tiempo extremadamente corto.

En resumen, la duración esperada de la operación para las víctimas en Triage A es extremadamente corta, generalmente dentro de los primeros minutos después de ser clasificadas. La rapidez en la intervención es crucial para estabilizar a las víctimas y prevenir complicaciones fatales, aumentando así sus posibilidades de supervivencia.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.030797101449275364, recall=0.85,
fmeasure=0.05944055944055944)
ROUGE-2 score: Score(precision=0.02722323049001815,
recall=0.7894736842105263, fmeasure=0.05263157894736842)
ROUGE-L score: Score(precision=0.030797101449275364, recall=0.85,
fmeasure=0.05944055944055944)
BERTScore - Precision: 0.5713352560997009, Recall: 0.74625563621521, F1:
0.647184431552887
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.031 indica que solo el 3.1% de las palabras generadas son relevantes en comparación con el texto de referencia.

- **Recall:** 0.85 sugiere que el sistema recupera el 85% de las palabras relevantes, lo que indica un **recall muy alto**.
 - **F1-Score:** 0.059 refleja un **desequilibrio significativo** entre precisión y recall, lo que indica que, aunque el sistema recupera muchas palabras relevantes, genera mucho contenido irrelevante.
2. **ROUGE-2 Score:**
- **Precisión:** 0.027 indica que solo el 2.7% de los bigrams generados son correctos en comparación con el texto de referencia.
 - **Recall:** 0.789 muestra que el sistema recupera el 78.9% de los bigrams relevantes.
 - **F1-Score:** 0.053 refleja un **rendimiento bajo** en la captura de secuencias significativas, con un **recall alto**, pero una **precisión extremadamente baja**.
3. **ROUGE-L Score:**
- **Precisión:** 0.031 sugiere que solo el 3.1% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.85 indica que el sistema recupera el 85% de la secuencia relevante más larga.
 - **F1-Score:** 0.059 muestra un **rendimiento bajo** en la captura de secuencias estructuradas, con **mucho más recuperación que precisión**.
4. **BERTScore:**
- **Precisión:** 0.571 indica que el 57.1% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.746 sugiere que el sistema captura el 74.6% del contenido semántico relevante.
 - **F1-Score:** 0.647 refleja un **rendimiento moderado** en términos semánticos, con un **buen recall** pero una **precisión limitada**.

Conclusión: Las métricas **ROUGE** indican que el sistema tiene un **recall muy alto**, pero una **precisión extremadamente baja**, lo que sugiere que genera una gran cantidad de contenido irrelevante. Aunque el sistema es eficaz en la **recuperación de contenido relevante**, genera muchos errores en la precisión, especialmente en la captura de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** refleja un rendimiento semántico **moderado**, lo que indica que el sistema capta bien el significado general, pero debe mejorar considerablemente en **precisión** para generar respuestas más relevantes y coherentes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

En la categoría de Triage A la duración esperada de la operación es menos de 12 horas.

Respuesta generada con RAG Básico:

En la categoría de Triage A, la duración esperada de la operación es menos de 12 horas.

Métricas:

```
ROUGE-1 score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
ROUGE-2 score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
ROUGE-L score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
BERTScore - Precision: 0.9313477277755737, Recall: 0.9457674026489258, F1:
0.9385021924972534
```

Análisis de las Métricas

1. ROUGE-1, ROUGE-2, y ROUGE-L:

- **Precisión, Recall y F1-Score:** Todos los valores son 1.0, lo que indica que el sistema ha generado una **coincidencia perfecta** con el texto de referencia, tanto en términos de palabras individuales (ROUGE-1), bigrams (ROUGE-2), como en la estructura general de las secuencias (ROUGE-L).

2. BERTScore:

- **Precisión:** 0.931 indica que el 93.1% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.946 sugiere que el sistema captura el 94.6% del contenido semántico relevante.
- **F1-Score:** 0.939 refleja un **rendimiento semántico excelente**, con un equilibrio muy fuerte entre precisión y recall.

Conclusión: Las métricas **ROUGE** muestran que el sistema ha logrado una **correspondencia perfecta** con el texto de referencia, lo que sugiere que las respuestas generadas coinciden completamente con las esperadas, tanto en términos de palabras como de secuencias. El **BERTScore** también indica un rendimiento semántico **altamente preciso**, aunque con un ligero margen de mejora en el recall. En conjunto, el sistema es **extremadamente preciso y eficaz** tanto en la literalidad como en el significado.

Respuesta generada con RAG con Haystack:

En la categoría de Triage A, la duración esperada de la operación es menos de 12 horas.

Métricas:

```
ROUGE-1 score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
ROUGE-2 score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
ROUGE-L score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
BERTScore - Precision: 0.9313477277755737, Recall: 0.9457674026489258, F1:
0.9385021924972534
```

Análisis de las Métricas.

1. ROUGE-1, ROUGE-2, y ROUGE-L:

- **Precisión, Recall y F1-Score:** Todos los valores son 1.0, lo que indica que el sistema ha generado una coincidencia **exacta** con el texto de referencia. Esto

significa que las palabras, secuencias de palabras (bigrams) y las estructuras de secuencias coinciden perfectamente con lo esperado.

2. BERTScore:

- **Precisión:** 0.931 indica que el 93.1% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.946 sugiere que el sistema captura el 94.6% del contenido semántico relevante.
- **F1-Score:** 0.939 refleja un **rendimiento semántico excepcional**, con un equilibrio muy fuerte entre precisión y recall.

Conclusión: El sistema ha alcanzado una **coincidencia total** con el texto de referencia según las métricas **ROUGE**, lo que indica una correspondencia literal perfecta en términos de contenido y estructura. El **BERTScore** también sugiere que el sistema es **altamente preciso en términos semánticos**, lo que significa que no solo genera respuestas literales correctas, sino que también mantiene una alta relevancia y coherencia semántica.

Respuesta generada con RAG con LLaIndex

La duración esperada de la operación en la categoría de Triage A es menos de 12 horas.

Métricas

```
ROUGE-1 score: Score(precision=1.0, recall=1.0, fmeasure=1.0)
ROUGE-2 score: Score(precision=0.8947368421052632,
recall=0.8947368421052632, fmeasure=0.8947368421052632)
ROUGE-L score: Score(precision=0.65, recall=0.65, fmeasure=0.65)
BERTScore - Precision: 0.9131824970245361, Recall: 0.9175553321838379, F1:
0.9153636693954468
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión, Recall y F1-Score:** 1.0 en todos los casos, lo que indica una **coincidencia perfecta** con el texto de referencia en términos de palabras individuales. Esto significa que todas las palabras generadas coinciden completamente con el texto de referencia.

2. ROUGE-2 Score:

- **Precisión, Recall y F1-Score:** 0.895 muestra que el sistema logra una **coincidencia muy alta** en los bigrams, pero no perfecta. Esto sugiere que aunque la mayoría de las secuencias de dos palabras son correctas, aún hay margen de mejora.

3. ROUGE-L Score:

- **Precisión, Recall y F1-Score:** 0.65 refleja un **rendimiento moderado** en la captura de la estructura general del texto y las secuencias más largas. Si bien el sistema genera palabras relevantes, puede haber problemas con la longitud o el orden de las secuencias.

4. BERTScore:

- **Precisión:** 0.913 indica que el 91.3% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.918 sugiere que el sistema captura el 91.8% del contenido semántico relevante.
- **F1-Score:** 0.915 refleja un **rendimiento semántico fuerte** y un **buen equilibrio** entre precisión y recall.

Conclusión: Las métricas **ROUGE-1** muestran que el sistema ha alcanzado una **coincidencia perfecta en palabras individuales**, pero **ROUGE-2** y **ROUGE-L** indican que el rendimiento en términos de secuencias y estructura del texto es bueno pero no perfecto, con un **rendimiento moderado** en la estructura general (ROUGE-L). El **BERTScore** sugiere que el sistema es **semánticamente sólido**, con una alta relevancia semántica en las respuestas generadas, pero puede mejorar en la **estructura y orden de las secuencias**.

Pregunta 9: ¿Qué es un sistema de Marcaje de INSARAG?

Respuesta generada por LECHAT


Search...

Today

Como se compone el código d...

CP

Qué es un sistema de Marcaje de INSARAG?

 El sistema de marcaje de INSARAG (Grupo Asesor Internacional de Búsqueda y Rescate) es una serie de símbolos y señales estandarizados utilizados para marcar y clasificar estructuras y áreas afectadas durante operaciones de búsqueda y rescate en situaciones de desastre. Este sistema es fundamental para la coordinación y comunicación efectiva entre los equipos de rescate internacionales, asegurando que todos los involucrados comprendan y sigan los mismos protocolos.

Componentes del Sistema de Marcaje de INSARAG

1. Marcaje de Estructuras:

- » **Evaluación Estructural:** Los equipos de rescate utilizan símbolos específicos para marcar estructuras que han sido evaluadas. Estos símbolos indican si una estructura ha sido inspeccionada y si es segura para entrar.
- » **Estado de la Estructura:** Se utilizan diferentes símbolos para indicar si una estructura es segura, parcialmente segura o insegura. Esto ayuda a los equipos de rescate a tomar decisiones informadas sobre dónde y cómo operar.

2. Marcaje de Áreas de Búsqueda

- » **Áreas Inspeccionadas:** Se utilizan símbolos para marcar áreas que han sido inspeccionadas y donde se han realizado operaciones de búsqueda y rescate.
- » **Áreas Sin Inspeccionar:** Se utilizan diferentes símbolos para indicar áreas que aún no han sido inspeccionadas, ayudando a los equipos de rescate a priorizar sus esfuerzos.

3. Marcaje de Víctimas:

- » **Víctimas Encontradas:** Se utilizan símbolos para marcar la ubicación de víctimas encontradas, ya sean vivas o fallecidas.
- » **Víctimas Rescatadas:** Se utilizan símbolos para indicar que una víctima ha sido rescatada y evacuada de la zona.

4. Marcaje de Peligros:

- » **Peligros Identificados:** Se utilizan símbolos para marcar peligros específicos, como fugas de gas, líneas eléctricas caídas, estructuras inestables, etc.
- » **Peligros Mitigados:** Se utilizan símbolos para indicar que un peligro ha sido identificado y mitigado, permitiendo operaciones seguras en el área.

Propósito del Sistema de Marcaje de INSARAG

El propósito del sistema de marcaje de INSARAG es facilitar la coordinación y comunicación entre los equipos de rescate internacionales, asegurando que todos los involucrados comprendan y sigan los mismos protocolos. Este sistema ayuda a:

1. **Mejorar la Eficiencia:** Al proporcionar una forma estandarizada de marcar y clasificar estructuras y áreas, el sistema de marcaje de INSARAG mejora la eficiencia de las operaciones de búsqueda y rescate.
2. **Reducir la Duplicación de Esfuerzos:** Al indicar claramente qué áreas han sido inspeccionadas y cuáles no, el sistema de marcaje ayuda a evitar la duplicación de esfuerzos y a priorizar las áreas que aún necesitan ser inspeccionadas.
3. **Aumentar la Seguridad:** Al marcar peligros específicos y el estado de las estructuras, el sistema de marcaje ayuda a aumentar la seguridad de los equipos de rescate y de las víctimas.
4. **Facilitar la Coordinación:** Al proporcionar una forma estandarizada de comunicación, el sistema de marcaje facilita la coordinación entre diferentes equipos de rescate, incluso aquellos que hablan diferentes idiomas.

Ejemplos de Símbolos de Marcaje

- **Círculo con una "X":** Indica que una estructura ha sido inspeccionada y no se encontraron víctimas.
- **Círculo con un "0":** Indica que una estructura ha sido inspeccionada y se encontraron víctimas vivas.
- **Círculo con un "+":** Indica que una estructura ha sido inspeccionada y se encontraron víctimas fallecidas.
- **Triángulo:** Indica un peligro específico, como una fuga de gas o una línea eléctrica caída.

En resumen, el sistema de marcaje de INSARAG es una herramienta esencial para la coordinación y comunicación efectiva durante operaciones de búsqueda y rescate en situaciones de desastre. Proporciona una forma estandarizada de marcar y clasificar estructuras y áreas, mejorando la eficiencia, reduciendo la duplicación de esfuerzos, aumentando la seguridad y facilitando la coordinación entre diferentes equipos de rescate.

Transcripción.

El sistema de marcaje de INSARAG (Grupo Asesor Internacional de Búsqueda y Rescate) es una serie de símbolos y señales estandarizados utilizados para marcar y clasificar estructuras y áreas afectadas durante operaciones de búsqueda y rescate en situaciones de desastre. Este sistema es fundamental para la coordinación y comunicación efectiva entre los equipos de rescate internacionales, asegurando que todos los involucrados comprendan y sigan los mismos protocolos.

Componentes del Sistema de Marcaje de INSARAG

1. Marcaje de Estructuras:

- **Evaluación Estructural:** Los equipos de rescate utilizan símbolos específicos para marcar estructuras que han sido evaluadas. Estos símbolos indican si una estructura ha sido inspeccionada y si es segura para entrar.
- **Estado de la Estructura:** Se utilizan diferentes símbolos para indicar si una estructura es segura, parcialmente segura o insegura. Esto ayuda a los equipos de rescate a tomar decisiones informadas sobre dónde y cómo operar.

2. Marcaje de Áreas de Búsqueda:

- **Áreas Inspeccionadas:** Se utilizan símbolos para marcar áreas que han sido inspeccionadas y donde se han realizado operaciones de búsqueda y rescate.
- **Áreas Sin Inspeccionar:** Se utilizan diferentes símbolos para indicar áreas que aún no han sido inspeccionadas, ayudando a los equipos de rescate a priorizar sus esfuerzos.

3. Marcaje de Víctimas:

- Víctimas Encontradas: Se utilizan símbolos para marcar la ubicación de víctimas encontradas, ya sean vivas o fallecidas.
- Víctimas Rescatadas: Se utilizan símbolos para indicar que una víctima ha sido rescatada y evacuada de la zona.

4. Marcaje de Peligros:

- Peligros Identificados: Se utilizan símbolos para marcar peligros específicos, como fugas de gas, líneas eléctricas caídas, estructuras inestables, etc.
- Peligros Mitigados: Se utilizan símbolos para indicar que un peligro ha sido identificado y mitigado, permitiendo operaciones seguras en el área.

Propósito del Sistema de Marcaje de INSARAG

El propósito del sistema de marcaje de INSARAG es facilitar la coordinación y comunicación entre los equipos de rescate internacionales, asegurando que todos los involucrados comprendan y sigan los mismos protocolos. Este sistema ayuda a:

1. Mejorar la Eficiencia: Al proporcionar una forma estandarizada de marcar y clasificar estructuras y áreas, el sistema de marcaje de INSARAG mejora la eficiencia de las operaciones de búsqueda y rescate.
2. Reducir la Duplicación de Esfuerzos: Al indicar claramente qué áreas han sido inspeccionadas y cuáles no, el sistema de marcaje ayuda a evitar la duplicación de esfuerzos y a priorizar las áreas que aún necesitan ser inspeccionadas.
3. Aumentar la Seguridad: Al marcar peligros específicos y el estado de las estructuras, el sistema de marcaje ayuda a aumentar la seguridad de los equipos de rescate y de las víctimas.
4. Facilitar la Coordinación: Al proporcionar una forma estandarizada de comunicación, el sistema de marcaje facilita la coordinación entre diferentes equipos de rescate, incluso aquellos que hablan diferentes idiomas.

Ejemplos de Símbolos de Marcaje

- Círculo con una "X": Indica que una estructura ha sido inspeccionada y no se encontraron víctimas.
- Círculo con un "O": Indica que una estructura ha sido inspeccionada y se encontraron víctimas vivas.
- Círculo con un "+": Indica que una estructura ha sido inspeccionada y se encontraron víctimas fallecidas.
- Triángulo: Indica un peligro específico, como una fuga de gas o una línea eléctrica caída.

En resumen, el sistema de marcaje de INSARAG es una herramienta esencial para la coordinación y comunicación efectiva durante operaciones de búsqueda y rescate en situaciones de desastre. Proporciona una forma estandarizada de marcar y clasificar estructuras y áreas, mejorando la eficiencia, reduciendo la duplicación de esfuerzos, aumentando la seguridad y facilitando la coordinación entre diferentes equipos de rescate.

Métricas sin RAG.

ROUGE-1 score: Score(precision=0.19440124416796267,
recall=0.5681818181818182, fmeasure=0.28968713789107764)
ROUGE-2 score: Score(precision=0.07165109034267912,
recall=0.2100456621004566, fmeasure=0.10685249709639953)
ROUGE-L score: Score(precision=0.1150855365474339,
recall=0.33636363636363636, fmeasure=0.17149478563151796)
BERTScore - Precision: 0.6576672196388245, Recall: 0.6875374913215637, F1:
0.6722707152366638

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.194 sugiere que solo el 19.4% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.568 muestra que el sistema recupera el 56.8% de las palabras relevantes, lo que indica un **recall moderado**.
- **F1-Score:** 0.290 refleja un **rendimiento bajo**, con un desequilibrio entre precisión y recall.

2. ROUGE-2 Score:

- **Precisión:** 0.072 indica que solo el 7.2% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.210 muestra que el sistema recupera el 21% de los bigrams relevantes.
- **F1-Score:** 0.107 refleja un **rendimiento bajo** en la captura de secuencias significativas, con una **precisión limitada** y un **recall bajo**.

3. ROUGE-L Score:

- **Precisión:** 0.115 sugiere que solo el 11.5% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.336 indica que el sistema recupera el 33.6% de la secuencia relevante más larga.
- **F1-Score:** 0.171 refleja un **rendimiento bajo** en la captura de secuencias estructuradas, con un **recall moderado** pero una **precisión baja**.

4. BERTScore:

- **Precisión:** 0.658 indica que el 65.8% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.688 sugiere que el sistema captura el 68.8% del contenido semántico relevante.
- **F1-Score:** 0.672 refleja un **rendimiento semántico moderado**, con un equilibrio razonable entre precisión y recall.

Conclusión: Las métricas **ROUGE** muestran un rendimiento **limitado en precisión**, con un **recall moderado** pero mucho contenido irrelevante generado. El sistema tiene dificultades para capturar secuencias significativas, como lo indican los **bajos valores de ROUGE-2 y ROUGE-L**. El **BERTScore** sugiere que el sistema es **moderadamente sólido en términos semánticos**, capturando bien el significado general, pero necesita mejorar en **precisión literal** y en la **estructura de las secuencias** para generar respuestas más coherentes y relevantes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

Los sistemas de Marcaje son una herramienta esencial usada en las operaciones USAR para presentar y compartir información clave entre equipos de rescate y demás personal de campo. También deben ser un mecanismo para fortalecer la coordinación y minimizar la duplicación. Para maximizar el valor de usar un sistema de Marcaje, en caso de que sea necesario identificar y usar universalmente un solo método común.

Para que este método sea efectivo, debe ser usado por todos los rescatistas, ser simple de aplicar, simple de entender, eficiente en el uso de recursos y tiempos, comunicar la información efectivamente y ser consistentemente aplicado.

El Sistema de Marcaje de INSARAG se empeña en alcanzar estos objetivos y consta de tres elementos principales de Marcaje: Marcaje Triage, Marcaje de Víctimas y Marcaje de Despeje Rápido (RCM). Estos componentes generan una serie detallada de presentaciones visuales que capturan información crítica para informar el conocimiento de la situación y apoyar la planeación y la coordinación.

El Sistema de Marcaje de INSARAG es usado por equipos como el sistema de Marcaje por defecto en ausencia de un sistema nacional en países en los cuales se presenten operaciones. El uso del sistema de Marcaje es determinado por LEMA y UCC.

Respuesta generada con RAG Básico:

Un sistema de Marcaje de INSARAG es un conjunto de procedimientos y métodos utilizados para facilitar la comunicación y coordinación entre equipos de rescate y otros personal de campo durante operaciones de emergencia. Este sistema está diseñado para ser simple de aplicar y entender, eficiente en el uso de recursos y tiempos, y para comunicar información crítica de manera efectiva y consistente.

El Sistema de Marcaje de INSARAG consta de tres elementos principales: Marcaje Triage, Marcaje de Víctimas y Marcaje de Despeje Rápido (RCM). Estos componentes generan una serie de presentaciones visuales detalladas que capturan información crucial para informar el conocimiento de la situación y apoyar la planeación y la coordinación.

El sistema es utilizado por equipos de rescate como el sistema de Marcaje por defecto en ausencia de un sistema nacional en los países donde se realicen operaciones. Su uso es determinado por LEMA y UCC. Se recomienda a los países que adopten el Sistema de Marcaje de INSARAG como su estándar nacional para facilitar la intervención de equipos internacionales en tiempos de crisis. El sistema está diseñado para complementar los sistemas nacionales y no para competir con ellos.

Métricas:

ROUGE-1 score: Score(precision=0.748792270531401,
recall=0.7045454545454546, fmeasure=0.7259953161592506)

```
ROUGE-2 score: Score(precision=0.49514563106796117,
recall=0.4657534246575342, fmeasure=0.48000000000000004)
ROUGE-L score: Score(precision=0.5652173913043478,
recall=0.5318181818181819, fmeasure=0.5480093676814988)
BERTScore - Precision: 0.8445764183998108, Recall: 0.8230332136154175, F1:
0.8336656093597412
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.749 sugiere que el 74.9% de las palabras generadas son relevantes.
- **Recall:** 0.705 indica que el sistema recupera el 70.5% de las palabras relevantes.
- **F1-Score:** 0.726 refleja un **buen equilibrio** entre precisión y recall, lo que indica que el sistema genera una cantidad significativa de contenido relevante.

2. ROUGE-2 Score:

- **Precisión:** 0.495 indica que el 49.5% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.466 sugiere que el sistema recupera el 46.6% de los bigrams relevantes.
- **F1-Score:** 0.480 refleja un **rendimiento moderado** en la captura de secuencias de dos palabras, con una precisión y recall aceptables, pero con margen de mejora.

3. ROUGE-L Score:

- **Precisión:** 0.565 sugiere que el 56.5% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.532 indica que el sistema recupera el 53.2% de la secuencia relevante más larga.
- **F1-Score:** 0.548 refleja un **rendimiento razonable** en la captura de secuencias estructuradas, con un **buen balance** entre precisión y recall.

4. BERTScore:

- **Precisión:** 0.845 indica que el 84.5% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.823 sugiere que el sistema captura el 82.3% del contenido semántico relevante.
- **F1-Score:** 0.834 refleja un **rendimiento semántico sólido**, con un buen equilibrio entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran un rendimiento **bueno** en términos de palabras individuales (ROUGE-1), aunque el sistema podría mejorar en la captura de **secuencias significativas** (ROUGE-2). El **BERTScore** indica que el sistema es **semánticamente sólido**, capturando el significado general de manera efectiva. En resumen, el sistema es eficaz en generar contenido relevante.

Respuesta generada con RAG con Haystack:

El Sistema de Marcaje de INSARAG es una herramienta esencial utilizada en operaciones de búsqueda y rescate urbano (USAR) para presentar y compartir información clave entre equipos de rescate y demás personal de campo. Este sistema tiene como objetivo fortalecer la coordinación y minimizar la

duplicación de esfuerzos. Para ser efectivo, debe ser utilizado por todos los rescatistas, ser simple de aplicar y entender, eficiente en el uso de recursos y tiempos, y comunicar la información de manera efectiva y consistente.

El Sistema de Marcaje de INSARAG consta de tres elementos principales: Marcaje Triage, Marcaje de Víctimas y Marcaje de Despeje Rápido (RCM). Estos componentes generan una serie detallada de presentaciones visuales que capturan información crítica para informar el conocimiento de la situación y apoyar la planeación y la coordinación.

Este sistema es utilizado como el sistema de Marcaje por defecto en ausencia de un sistema nacional en países donde se presenten operaciones. El uso del sistema de Marcaje es determinado por LEMA y UCC. Se recomienda a los países que utilicen el Sistema de Marcaje de INSARAG como su estándar nacional, ya que está diseñado para complementar los sistemas nacionales y no para competir con ellos.

Métricas:

ROUGE-1 score: Score(precision=0.8130841121495327, recall=0.7909090909090909, fmeasure=0.8018433179723501)
ROUGE-2 score: Score(precision=0.6525821596244131, recall=0.634703196347032, fmeasure=0.6435185185185186)
ROUGE-L score: Score(precision=0.705607476635514, recall=0.6863636363636364, fmeasure=0.6958525345622121)
BERTScore - Precision: 0.8900979161262512, Recall: 0.8846827745437622, F1: 0.8873820900917053

Análisis de las Métricas.

1. ROUGE-1 Score:

- **Precisión:** 0.813 indica que el 81.3% de las palabras generadas son relevantes.
- **Recall:** 0.791 muestra que el sistema recupera el 79.1% de las palabras relevantes.
- **F1-Score:** 0.802 refleja un **buen equilibrio** entre precisión y recall, indicando que el sistema es eficaz en generar palabras relevantes con una precisión alta.

2. ROUGE-2 Score:

- **Precisión:** 0.653 sugiere que el 65.3% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.635 indica que el sistema recupera el 63.5% de los bigrams relevantes.
- **F1-Score:** 0.644 refleja un **rendimiento sólido** en la captura de secuencias de palabras, aunque con margen para mejorar la precisión en las secuencias más largas.

3. ROUGE-L Score:

- **Precisión:** 0.706 indica que el 70.6% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.686 sugiere que el sistema recupera el 68.6% de la secuencia relevante más larga.

- **F1-Score:** 0.696 refleja un **buen rendimiento estructural**, lo que sugiere que el sistema genera secuencias coherentes.
4. **BERTScore:**
- **Precisión:** 0.890 indica que el 89.0% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.885 sugiere que el sistema captura el 88.5% del contenido semántico relevante.
 - **F1-Score:** 0.887 refleja un **rendimiento semántico excelente**, con un equilibrio muy sólido entre precisión y recall.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene un **rendimiento sólido** tanto en términos de palabras individuales (ROUGE-1) como en secuencias de palabras (ROUGE-2 y ROUGE-L). El **BERTScore** refleja que el sistema es **altamente efectivo en términos semánticos**, capturando la mayor parte del contenido relevante con una alta precisión. En general, el sistema es muy eficiente, aunque hay un pequeño margen de mejora en la **precisión de las secuencias más largas** para perfeccionar aún más la coherencia estructural.

Respuesta generada con RAG con LlamaIndex

Un sistema de Marcaje de INSARAG es una herramienta esencial utilizada en operaciones de búsqueda y rescate urbano (USAR) para presentar y compartir información clave entre equipos de rescate y otros personal de campo. Su objetivo es fortalecer la coordinación y minimizar la duplicación de esfuerzos. Este sistema consta de tres elementos principales: Marcaje Triage, Marcaje de Víctimas y Marcaje de Despeje Rápido (RCM). Estos componentes generan una serie detallada de presentaciones visuales que capturan información crítica para informar el conocimiento de la situación y apoyar la planeación y la coordinación.

Métricas

```
ROUGE-1 score: Score(precision=0.9215686274509803,
recall=0.42727272727272725, fmeasure=0.5838509316770187)
ROUGE-2 score: Score(precision=0.7722722727227273,
recall=0.3561643835616438, fmeasure=0.48750000000000004)
ROUGE-L score: Score(precision=0.8431372549019608,
recall=0.39090909090909093, fmeasure=0.5341614906832298)
BERTScore - Precision: 0.9171077609062195, Recall: 0.7956551909446716, F1:
0.8520753979682922
```

Análisis de las Métricas

1. **ROUGE-1 Score:**
- **Precisión:** 0.922 indica que el 92.2% de las palabras generadas son relevantes.
 - **Recall:** 0.427 muestra que el sistema recupera el 42.7% de las palabras relevantes.

- **F1-Score:** 0.584 refleja un **desequilibrio entre precisión y recall**, con una alta precisión pero un recall limitado, lo que sugiere que el sistema genera palabras muy relevantes pero no recupera todo el contenido esperado.
- 2. **ROUGE-2 Score:**
 - **Precisión:** 0.772 sugiere que el 77.2% de los bigrams generados son correctos.
 - **Recall:** 0.356 muestra que el sistema recupera el 35.6% de los bigrams relevantes.
 - **F1-Score:** 0.488 refleja un **rendimiento moderado** en la captura de secuencias, con una buena precisión pero un recall limitado.
- 3. **ROUGE-L Score:**
 - **Precisión:** 0.843 indica que el 84.3% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
 - **Recall:** 0.391 sugiere que el sistema recupera el 39.1% de la secuencia relevante más larga.
 - **F1-Score:** 0.534 refleja un **rendimiento moderado en la captura de secuencias estructuradas**, con un mejor rendimiento en precisión que en recall.
- 4. **BERTScore:**
 - **Precisión:** 0.917 indica que el 91.7% de las palabras generadas son semánticamente relevantes.
 - **Recall:** 0.796 sugiere que el sistema captura el 79.6% del contenido semántico relevante.
 - **F1-Score:** 0.852 refleja un **rendimiento semántico sólido**, con un buen equilibrio entre precisión y recall.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene una **alta precisión**, pero el **recall es limitado**, lo que sugiere que el sistema genera contenido muy relevante pero no recupera todo el contenido disponible, especialmente en términos de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** indica un **rendimiento semántico fuerte**, capturando bien el significado general, aunque con espacio para mejorar en **recuperación de contenido**.

Pregunta 10: ¿Cuáles son las consideraciones estratégicas que se deben tener en cuenta en las operaciones con materiales peligrosos?

Respuesta generada por LECHAT

Search...

Today

Como se compone el código d...

Como se compone el código d...

Como se compone el código d...

CP

cuáles son las consideraciones estratégicas que se deben tener en cuenta en las operaciones con materiales peligrosos?

M

Las operaciones con materiales peligrosos (HAZMAT, por sus siglas en inglés) requieren una planificación y ejecución cuidadosas para minimizar riesgos y asegurar la seguridad de los rescatistas, las víctimas y el entorno. A continuación se presentan algunas consideraciones estratégicas clave que se deben tener en cuenta en estas operaciones:

1. Evaluación y Reconocimiento Inicial

- Identificación del Material Peligroso: Identificar el tipo de material peligroso involucrado y sus propiedades (tóxico, inflamable, corrosivo, etc.).
- Evaluación del Riesgo: Evaluar el riesgo potencial para los rescatistas y las víctimas. Incluyendo la posibilidad de exposición, dispersión y reacciones químicas.

2. Seguridad del Personal

- Equipos de Protección Personal (EPP): Asegurarse de que los rescatistas utilicen el equipo de protección personal adecuado, como trajes de protección química, guantes, gafas y respiradores.
- Capacitación: Asegurarse de que el personal esté debidamente capacitado en el manejo de materiales peligrosos y en los procedimientos de seguridad.

3. Establecimiento de Zonas de Seguridad

- Zona Caliente: Área donde se encuentra el material peligroso y donde el riesgo de exposición es más alto.
- Zona Templada: Área de descontaminación donde los rescatistas y las víctimas pueden ser descontaminados antes de salir de la zona caliente.
- Zona Fría: Área segura donde se coordinan las operaciones y se proporciona atención médica.

4. Descontaminación

- Procedimientos de Descontaminación: Establecer procedimientos claros para la descontaminación de personal, equipos y víctimas antes de salir de la zona caliente.
- Estaciones de Descontaminación: Configurar estaciones de descontaminación con los recursos necesarios, como agua, jabón, soluciones descontaminantes y contenedores para residuos contaminados.

5. Coordinación y Comunicación

- Comando de Incidentes: Establecer un sistema de comando de incidentes para coordinar las operaciones y asegurar una comunicación efectiva entre todos los equipos involucrados.
- Comunicación Continua: Mantener una comunicación continua entre los equipos de rescate, los coordinadores de la respuesta y otras agencias involucradas.

6. Planificación y Logística

- Recursos y Equipos: Asegurarse de que se disponga de los recursos y equipos necesarios, incluyendo equipos de protección personal, equipos de detección y monitoreo, y suministros de descontaminación.
- Plan de Respuesta: Desarrollar un plan de respuesta detallado que incluya procedimientos para la identificación, contención, descontaminación y eliminación de materiales peligrosos.

7. Contención y Mitigación

- Contención del Material Peligroso: Utilizar métodos adecuados para contener el material peligroso y prevenir su dispersión, como barreras físicas, absorbentes y sellantes.
- Mitigación de Riesgos: Implementar medidas para mitigar los riesgos asociados con el material

<p>Today</p> <p>Como se compone el código d...</p>		<p>peligroso, como la neutralización de sustancias químicas y la ventilación de áreas contaminadas.</p> <p>8. Atención Médica</p> <ul style="list-style-type: none">• Evaluación y Tratamiento Inicial: Proporcionar evaluación y tratamiento médico inicial a las víctimas expuestas a materiales peligrosos.• Transporte Seguro: Asegurar el transporte seguro de las víctimas a instalaciones médicas adecuadas para recibir atención adicional. <p>9. Documentación y Registro</p> <ul style="list-style-type: none">• Registro de Incidentes: Mantener un registro detallado de todos los incidentes, incluyendo la identificación del material peligroso, las acciones tomadas y los resultados.• Informes de Progreso: Proporcionar informes regulares sobre el progreso de las operaciones a los coordinadores de la respuesta y otras partes interesadas. <p>10. Evaluación Post-Incidente</p> <ul style="list-style-type: none">• Revisión de Procedimientos: Revisar los procedimientos y las acciones tomadas durante el incidente para identificar áreas de mejora.• Capacitación y Actualización: Utilizar las lecciones aprendidas para mejorar la capacitación y actualizar los procedimientos de respuesta a materiales peligrosos. <p>En resumen, las consideraciones estratégicas en las operaciones con materiales peligrosos incluyen la evaluación y reconocimiento inicial, la seguridad del personal, el establecimiento de zonas de seguridad, la descontaminación, la coordinación y comunicación, la planificación y logística, la contención y mitigación, la atención médica, la documentación y registro, y la evaluación post-incidente. Estas consideraciones son esenciales para asegurar una respuesta efectiva y segura en situaciones que involucren</p>
--	--	--

Transcripción.

Las operaciones con materiales peligrosos (HAZMAT, por sus siglas en inglés) requieren una planificación y ejecución cuidadosas para minimizar riesgos y asegurar la seguridad de los rescatistas, las víctimas y el entorno. A continuación se presentan algunas consideraciones estratégicas clave que se deben tener en cuenta en estas operaciones:

1. Evaluación y Reconocimiento Inicial

- **Identificación del Material Peligroso:** Identificar el tipo de material peligroso involucrado y sus propiedades (tóxico, inflamable, corrosivo, etc.).
- **Evaluación del Riesgo:** Evaluar el riesgo potencial para los rescatistas y las víctimas, incluyendo la posibilidad de exposición, dispersión y reacciones químicas.

2. Seguridad del Personal

- **Equipos de Protección Personal (EPP):** Asegurarse de que los rescatistas utilicen el equipo de protección personal adecuado, como trajes de protección química, guantes, gafas y respiradores.
- **Capacitación:** Asegurarse de que el personal esté debidamente capacitado en el manejo de materiales peligrosos y en los procedimientos de seguridad.

3. Establecimiento de Zonas de Seguridad

- **Zona Caliente:** Área donde se encuentra el material peligroso y donde el riesgo de exposición es más alto.
- **Zona Templada:** Área de descontaminación donde los rescatistas y las víctimas pueden ser descontaminados antes de salir de la zona caliente.
- **Zona Fría:** Área segura donde se coordinan las operaciones y se proporciona atención médica.

4. Descontaminación

- **Procedimientos de Descontaminación:** Establecer procedimientos claros para la descontaminación de personal, equipos y víctimas antes de salir de la zona caliente.
- **Estaciones de Descontaminación:** Configurar estaciones de descontaminación con los recursos necesarios, como agua, jabón, soluciones descontaminantes y contenedores para residuos contaminados.

5. Coordinación y Comunicación

- **Comando de Incidentes:** Establecer un sistema de comando de incidentes para coordinar las operaciones y asegurar una comunicación efectiva entre todos los equipos involucrados.
- **Comunicación Continua:** Mantener una comunicación continua entre los equipos de rescate, los coordinadores de la respuesta y otras agencias involucradas.

6. Planificación y Logística

- **Recursos y Equipos:** Asegurarse de que se disponga de los recursos y equipos necesarios, incluyendo equipos de protección personal, equipos de detección y monitoreo, y suministros de descontaminación.
- **Plan de Respuesta:** Desarrollar un plan de respuesta detallado que incluya procedimientos para la identificación, contención, descontaminación y eliminación de materiales peligrosos.

7. Contención y Mitigación

- **Contención del Material Peligroso:** Utilizar métodos adecuados para contener el material peligroso y prevenir su dispersión, como barreras físicas, absorbentes y sellantes.
- **Mitigación de Riesgos:** Implementar medidas para mitigar los riesgos asociados con el material peligroso, como la neutralización de sustancias químicas y la ventilación de áreas contaminadas.

8. Atención Médica

- **Evaluación y Tratamiento Inicial:** Proporcionar evaluación y tratamiento médico inicial a las víctimas expuestas a materiales peligrosos.
- **Transporte Seguro:** Asegurar el transporte seguro de las víctimas a instalaciones médicas adecuadas para recibir atención adicional.

9. Documentación y Registro

- **Registro de Incidentes:** Mantener un registro detallado de todos los incidentes, incluyendo la identificación del material peligroso, las acciones tomadas y los resultados.
- **Informes de Progreso:** Proporcionar informes regulares sobre el progreso de las operaciones a los coordinadores de la respuesta y otras partes interesadas.

10. Evaluación Post-Incidente

- **Revisión de Procedimientos:** Revisar los procedimientos y las acciones tomadas durante el incidente para identificar áreas de mejora.
- **Capacitación y Actualización:** Utilizar las lecciones aprendidas para mejorar la capacitación y actualizar los procedimientos de respuesta a materiales peligrosos.

En resumen, las consideraciones estratégicas en las operaciones con materiales peligrosos incluyen la evaluación y reconocimiento inicial, la seguridad del personal, el establecimiento de zonas de seguridad, la descontaminación, la coordinación y comunicación, la planificación y logística, la contención y mitigación, la atención médica, la documentación y registro, y la evaluación post-incidente. Estas consideraciones son esenciales para asegurar una respuesta efectiva y segura en situaciones que involucran.

Métricas sin RAG.

```
ROUGE-1 score: Score(precision=0.1111111111111111, recall=0.562962962962963, fmeasure=0.1855921855921856)
ROUGE-2 score: Score(precision=0.032210834553440704, recall=0.16417910447761194, fmeasure=0.05385556915544675)
ROUGE-L score: Score(precision=0.06432748538011696, recall=0.32592592592592595, fmeasure=0.10744810744810744)
BERTScore - Precision: 0.6620516180992126, Recall: 0.691678524017334, F1: 0.6765408515930176
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.111 indica que solo el 11.1% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.563 sugiere que el sistema recupera el 56.3% de las palabras relevantes, lo que indica un **recall moderado**.
- **F1-Score:** 0.186 refleja un **rendimiento bajo**, con un **desequilibrio entre precisión y recall**, lo que indica que el sistema genera mucho contenido irrelevante.

2. ROUGE-2 Score:

- **Precisión:** 0.032 indica que solo el 3.2% de los bigrams generados son correctos en comparación con el texto de referencia.
- **Recall:** 0.164 muestra que el sistema recupera el 16.4% de los bigrams relevantes.
- **F1-Score:** 0.054 refleja un **rendimiento bajo** en la captura de secuencias de palabras significativas, con una precisión muy limitada.

3. ROUGE-L Score:

- **Precisión:** 0.064 sugiere que solo el 6.4% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.326 indica que el sistema recupera el 32.6% de la secuencia relevante más larga.
- **F1-Score:** 0.107 refleja un **rendimiento bajo** en la captura de secuencias estructuradas, con una **precisión muy baja**.

4. BERTScore:

- **Precisión:** 0.662 indica que el 66.2% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.692 sugiere que el sistema captura el 69.2% del contenido semántico relevante.
- **F1-Score:** 0.677 refleja un **rendimiento semántico moderado**, con un equilibrio aceptable entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran un **recall moderado**, pero una **precisión extremadamente baja**, lo que indica que el sistema genera una gran cantidad de contenido irrelevante, especialmente en la captura de secuencias (ROUGE-2 y ROUGE-L). El **BERTScore** refleja un rendimiento **semántico moderado**, lo que sugiere que el sistema capta razonablemente bien el significado general, aunque necesita mejorar considerablemente en **precisión literal** y en la **estructura de las secuencias** para generar respuestas más coherentes y relevantes.

Respuesta esperada (en las tres implementaciones es la misma respuesta):

Se requiere que los Equipos USAR Livianos, Medianos y Pesados cuenten con el conocimiento inherente para reconocer un entorno peligroso, minimizando el riesgo de peligro, lesiones o muerte a sus miembros, la población afectada y el medioambiente. También se espera que los equipos puedan comunicarle sus hallazgos relacionados con contaminación a otros. como ya se indicó, un Equipo USAR debe:

Tener la capacidad de reconocer situaciones en las que se sospeche de contaminantes.

Tener la experticia técnica para ofrecerle asesoría a LEMA, UCC y otros actores.

Tener la capacidad de brindarles protección básica a los miembros de los equipos efectuando detección y monitoreo medioambiental.

Implementar procedimientos básicos de descontaminación.

Estar al tanto de las limitaciones del equipo al tratar con operaciones con materiales tóxicos.

Respuesta generada con RAG Básico:

Las consideraciones estratégicas que se deben tener en cuenta en las operaciones con materiales peligrosos son las siguientes:

1. ****Reconocer situaciones de contaminación**:** Los Equipos USAR deben tener la capacidad de reconocer situaciones en las que se sospeche de contaminantes.

2. ****Experticia técnica**:** Los equipos deben tener la experticia técnica para ofrecer asesoría a LEMA (Lugar de Emergencia Médica), UCC (Unidad de Coordinación Central) y otros actores.

3. ****Protección básica**:** Los equipos deben tener la capacidad de brindar protección básica a los miembros del equipo, efectuando detección y monitoreo medioambiental.

4. ****Procedimientos de descontaminación****: Los equipos deben implementar procedimientos básicos de descontaminación.

5. ****Limitaciones del equipo****: Los equipos deben estar al tanto de las limitaciones del equipo al tratar con operaciones con materiales tóxicos.

Estas consideraciones estratégicas son esenciales para minimizar el riesgo de peligro, lesiones o muerte a los miembros del equipo, la población afectada y el medioambiente.

Métricas:

```
ROUGE-1 score: Score(precision=0.6047904191616766,  
recall=0.7481481481481481, fmeasure=0.6688741721854304)  
ROUGE-2 score: Score(precision=0.4578313253012048,  
recall=0.5671641791044776, fmeasure=0.5066666666666666)  
ROUGE-L score: Score(precision=0.4550898203592814,  
recall=0.562962962962963, fmeasure=0.5033112582781457)  
BERTScore - Precision: 0.7569805383682251, Recall: 0.7983359098434448, F1:  
0.7771084308624268
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión**: 0.605 indica que el 60.5% de las palabras generadas son relevantes.
- **Recall**: 0.748 sugiere que el sistema recupera el 74.8% de las palabras relevantes, lo que indica un **recall sólido**.
- **F1-Score**: 0.669 refleja un **buen equilibrio** entre precisión y recall, mostrando que el sistema genera una cantidad significativa de contenido relevante.

2. ROUGE-2 Score:

- **Precisión**: 0.458 indica que el 45.8% de los bigrams generados coinciden con el texto de referencia.
- **Recall**: 0.567 sugiere que el sistema recupera el 56.7% de los bigrams relevantes.
- **F1-Score**: 0.507 refleja un **rendimiento moderado** en la captura de secuencias de palabras, con un balance aceptable entre precisión y recall, pero con margen de mejora en precisión.

3. ROUGE-L Score:

- **Precisión**: 0.455 indica que el 45.5% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall**: 0.563 muestra que el sistema recupera el 56.3% de la secuencia relevante más larga.
- **F1-Score**: 0.503 refleja un **rendimiento razonable** en la captura de secuencias estructuradas, aunque podría beneficiarse de una mayor precisión.

4. BERTScore:

- **Precisión**: 0.757 indica que el 75.7% de las palabras generadas son semánticamente relevantes.

- **Recall:** 0.798 sugiere que el sistema captura el 79.8% del contenido semántico relevante.
- **F1-Score:** 0.777 refleja un **rendimiento semántico fuerte**, con un buen equilibrio entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran un rendimiento **sólido** tanto en términos de palabras individuales (ROUGE-1) como en secuencias (ROUGE-2 y ROUGE-L), con un buen equilibrio entre precisión y recall. El **BERTScore** indica que el sistema es **semánticamente fuerte**, capturando bien el contenido relevante. En general, el sistema genera respuestas relevantes y coherentes, pero podría mejorar en la **precisión de secuencias** para aumentar la coherencia estructural y la relevancia en las secuencias más largas.

Respuesta generada con RAG con Haystack:

Las consideraciones estratégicas que se deben tener en cuenta en las operaciones con materiales peligrosos, según el documento proporcionado, son las siguientes:

1. ****Reconocimiento de situaciones de contaminación**:** Los equipos USAR deben tener la capacidad de reconocer situaciones en las que se sospeche de contaminantes.
2. ****Experticia técnica**:** Los equipos deben contar con la experticia técnica para ofrecer asesoría a LEMA (Local Emergency Management Authority), UCC (Unified Command Center) y otros actores.
3. ****Protección básica**:** Los equipos deben brindar protección básica a sus miembros efectuando detección y monitoreo medioambiental.
4. ****Procedimientos de descontaminación**:** Implementar procedimientos básicos de descontaminación.
5. ****Limitaciones del equipo**:** Estar al tanto de las limitaciones del equipo al tratar con operaciones con materiales tóxicos.

Estas consideraciones estratégicas son esenciales para minimizar el riesgo de peligro, lesiones o muerte a los miembros del equipo, la población afectada y el medioambiente. Además, los equipos deben ser capaces de comunicar sus hallazgos relacionados con la contaminación a otros actores involucrados.

Métricas:

```
ROUGE-1 score: Score(precision=0.5921787709497207,
recall=0.7851851851851852, fmeasure=0.6751592356687899)
ROUGE-2 score: Score(precision=0.43820224719101125,
recall=0.582089552238806, fmeasure=0.5)
ROUGE-L score: Score(precision=0.39106145251396646,
recall=0.5185185185185185, fmeasure=0.4458598726114649)
BERTScore - Precision: 0.7544669508934021, Recall: 0.8055382966995239, F1:
0.7791666984558105
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.592 sugiere que el 59.2% de las palabras generadas son relevantes.
- **Recall:** 0.785 indica que el sistema recupera el 78.5% de las palabras relevantes, lo que refleja un **recall alto**.
- **F1-Score:** 0.675 muestra un **buen equilibrio** entre precisión y recall, lo que sugiere que el sistema genera una cantidad significativa de contenido relevante con una alta recuperación.

2. ROUGE-2 Score:

- **Precisión:** 0.438 indica que el 43.8% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.582 sugiere que el sistema recupera el 58.2% de los bigrams relevantes.
- **F1-Score:** 0.500 refleja un **rendimiento moderado** en la captura de secuencias de dos palabras, con una **precisión aceptable** pero que podría mejorar.

3. ROUGE-L Score:

- **Precisión:** 0.391 sugiere que el 39.1% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.519 indica que el sistema recupera el 51.9% de la secuencia relevante más larga.
- **F1-Score:** 0.446 refleja un **rendimiento limitado** en la captura de secuencias estructuradas, lo que sugiere que el sistema podría mejorar la coherencia en la generación de secuencias más largas.

4. BERTScore:

- **Precisión:** 0.754 indica que el 75.4% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.806 sugiere que el sistema captura el 80.6% del contenido semántico relevante.
- **F1-Score:** 0.779 refleja un **rendimiento semántico fuerte**, con un buen equilibrio entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene un **buen rendimiento en términos de palabras individuales (ROUGE-1)**, pero hay **espacio para mejorar en la captura de secuencias significativas (ROUGE-2 y ROUGE-L)**. Aunque el sistema es bueno en **recuperar contenido relevante**, podría beneficiarse de una mayor precisión en la generación de secuencias estructuradas. El **BERTScore** indica que el sistema es **semánticamente fuerte**, capturando bien el significado general y con un buen equilibrio entre precisión y recall.

Respuesta generada con RAG con LlamaIndex

Las consideraciones estratégicas en las operaciones con materiales peligrosos incluyen la capacidad de reconocer situaciones en las que se sospeche de contaminantes, tener la experticia técnica para ofrecer asesoría a LEMA, UCC y otros actores, brindar protección básica a los miembros del equipo mediante detección y monitoreo medioambiental, implementar

procedimientos básicos de descontaminación y estar al tanto de las limitaciones del equipo al tratar con operaciones con materiales tóxicos.

Métricas

```
ROUGE-1 score: Score(precision=0.8333333333333334,  
recall=0.48148148148148145, fmeasure=0.6103286384976525)  
ROUGE-2 score: Score(precision=0.7272727272727273,  
recall=0.417910447761194, fmeasure=0.5308056872037914)  
ROUGE-L score: Score(precision=0.8205128205128205,  
recall=0.4740740740740741, fmeasure=0.6009389671361502)  
BERTScore - Precision: 0.8663592338562012, Recall: 0.772144079208374, F1:  
0.81654292345047
```

Análisis de las Métricas

1. ROUGE-1 Score:

- **Precisión:** 0.833 indica que el 83.3% de las palabras generadas son relevantes en comparación con el texto de referencia.
- **Recall:** 0.481 sugiere que el sistema recupera el 48.1% de las palabras relevantes.
- **F1-Score:** 0.610 refleja un **rendimiento moderado**, con una **alta precisión**, pero un **recall limitado**, lo que significa que el sistema genera contenido muy relevante, pero no recupera todo el contenido esperado.

2. ROUGE-2 Score:

- **Precisión:** 0.727 indica que el 72.7% de los bigrams generados coinciden con el texto de referencia.
- **Recall:** 0.418 muestra que el sistema recupera el 41.8% de los bigrams relevantes.
- **F1-Score:** 0.531 refleja un **rendimiento moderado** en la captura de secuencias significativas, con una **buena precisión** pero un **recall bajo**.

3. ROUGE-L Score:

- **Precisión:** 0.821 indica que el 82.1% de las coincidencias generadas pertenecen a la secuencia más larga de coincidencias.
- **Recall:** 0.474 sugiere que el sistema recupera el 47.4% de la secuencia relevante más larga.
- **F1-Score:** 0.601 refleja un **rendimiento sólido en secuencias estructuradas**, aunque el **recall** es más bajo que la precisión.

4. BERTScore:

- **Precisión:** 0.866 indica que el 86.6% de las palabras generadas son semánticamente relevantes.
- **Recall:** 0.772 sugiere que el sistema captura el 77.2% del contenido semántico relevante.
- **F1-Score:** 0.817 refleja un **rendimiento semántico fuerte**, con un **buen equilibrio** entre precisión y recall en términos de significado general.

Conclusión: Las métricas **ROUGE** muestran que el sistema tiene una **precisión alta** en la generación de contenido relevante, pero el **recall es limitado**, lo que significa que no captura todo el contenido esperado. El **BERTScore** refleja que el sistema es **semánticamente sólido**,

Anexo Emergencias y Ayuda Sanitaria.

TFM

Alumna: María Carolina Passarello.

capturando bien el significado general con una buena precisión. Para mejorar, sería ideal aumentar el **recall** sin comprometer la precisión, lo que ayudaría a generar respuestas más completas y coherentes.