

CONSOLIDADO DE OBSERVACIONES Y RESPUESTAS AL SONDEO DE MERCADO

Objeto: “Indexar la información documental generada por los proyectos de la transición energética y disponerla para búsquedas eficientes, utilizando nuevas tecnologías de la información y ciencia de la computación, mediante la adquisición e implementación de una solución tecnológica de indexación y búsqueda de información.”

PREGUNTAS PRESENTADAS POR – ARCETEC SOLUCIONES TECNOLÓGICAS

Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos

1. ¿Cuál es el propósito principal de la clasificación de estos documentos? ¿Es para la búsqueda eficiente de información, para la organización de una base de datos, o para otro propósito?

Respuesta: El propósito principal de la indexación de la información es la búsqueda, acceso y consulta eficiente de información, esto mediante una herramienta de indexación y búsqueda que integre nuevas tecnologías de la información y ciencia de la computación.

2. ¿Qué tipo de documentos se van a clasificar? ¿Son informes técnicos, artículos científicos, mapas, imágenes, datos tabulares u otros formatos?

Respuesta: Tal como se especifica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” y en la Actividad 2, la información a indexar corresponde a tipo documental la cual no requiere software especializado o petrotécnico para su visualización y manejo. En particular, esta información tiene contenido técnico y no técnico que encapsula los aspectos más destacados y cruciales de los contratos, convenios y proyectos gestionados por la VT.

“Dicha información puede corresponder a imágenes de campo (afloramientos rocosos, muestras de roca, núcleos, ripios, fotografías de secciones delgadas, puntos de agua, entre otros), información derivada de perforaciones y pozos (curvas de registros, ensayos y pruebas de pozos, geoquímica de rocas, hidrogeoquímica, análisis isotópicos, entre otros), perfiles e imágenes sísmicas, imágenes espectrales e hiperespectrales, imágenes de tomografía, DRX, FRX, SEM, espectroscopía Raman, secciones estructurales, mapas estructurales, columnas estratigráficas, bloques diagrama, batimetrías y geofísica marina, cartografía (geológica, geomorfológica, geofísica, temática, entre otras), datos de sensores remotos e imágenes satelitales, entre otros.”

“Es fundamental que durante la implementación se reconozca tanto la información documental técnica como la no técnica que pueda encapsular los aspectos más destacados cruciales de los contratos, convenios y proyectos gestionados por la VT. Esto abarca desde estudios previos, sondeos de mercado, comunicaciones internas, minutas, actas de inicio, actas de finalización, actas de liquidación, otrosí, actas de entregas, hasta productos en formatos no técnicos como informes finales, informes de actividades y supervisión. No obstante, es crucial señalar que, dada la disposición actual de la información, la cual mayormente carece de estructura, se torna imperativo identificar y ubicar la información definitiva y esencial almacenada en nuestro directorio colaborativo. Esto se debe a la

posibilidad de que existan versiones preliminares de cada uno de los documentos mencionados anteriormente.”

3. ¿Cuántas categorías o temas principales deseas incluir en la clasificación? ¿Se enfocará en áreas específicas de la geología, geofísica y geoespacial, como exploración de minerales, prospección petrolera, análisis de terremotos, cartografía, etc.?

Respuesta: Tal como se indica en la descripción de la necesidad, se requiere una solución tecnológica que permita búsquedas, acceso y consultas eficientes de información. Ahora bien, la manera en que se lleva a cabo la tipología, clasificación y catalogación documental, lo cual es fundamental en la indexación y búsqueda de información, no obedece a una manera en particular de hacerlo, ni a unas etiquetas y categorías específicas. Sin embargo, la propuesta de los interesados en materia de lo anteriormente mencionado debe responder eficientemente a los aspectos técnicos y la necesidad especificados en el sondeo de mercado.

4. ¿Cómo se van a adquirir y preprocesar los documentos? ¿Se utilizará un motor de búsqueda para recopilar documentos relevantes? ¿Necesitarás técnicas de limpieza y normalización de texto?

Respuesta: Tal como se describe en la descripción de la necesidad, en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” y en la Actividad 2, la información actualmente se encuentra en su mayoría no estructurada. lo anterior conlleva a que se requiera un preprocesamiento para identificar, clasificar y validar la veracidad, integridad, confiabilidad y consistencia de la información. Solo así se podrá recopilar la información correcta para la implementación de la herramienta de indexación y búsqueda. Es fundamental que los interesados propongan maneras eficientes e innovadoras para llevar a cabo este preprocesamiento.

5. ¿Cuáles serán las características relevantes de los documentos para la clasificación? ¿Palabras clave, temas recurrentes, estructura del documento, metadatos, etc.?

Las preguntas 3 y 5 están interconectadas.

Respuesta: Tal como se indica en la descripción de la necesidad, se requiere una solución tecnológica que permita búsquedas, acceso y consultas eficientes de información. Ahora bien, la manera en que se lleva a cabo la tipología, clasificación y catalogación documental, lo cual es fundamental en la indexación y búsqueda de información, no obedece a una manera en particular de hacerlo, ni a unas etiquetas y categorías específicas. Sin embargo, la propuesta de los interesados en materia de lo anteriormente mencionado debe responder eficientemente a los aspectos técnicos y la necesidad especificados en el sondeo de mercado.

6. ¿Qué modelo de lenguaje natural (LLM) planean utilizar? ¿Una red neuronal convolucional (CNN), una red neuronal recurrente (RNN), un transformer como BERT, o alguna otra arquitectura?

Respuesta: El Large Language Model (LLM) corresponde a una propiedad fundamental de la herramienta o solución que se busca adquirir e implementar. Es responsabilidad de los interesados proveer información sobre el LLM utilizado en la solución propuesta. Este mismo será considerado teniendo en cuenta que es una propiedad fundamental que determina las

capacidades que posee la solución y la cobertura de esta frente a las necesidades a las que se busca dar solución.

7. ¿Cómo se entrenará el modelo? ¿Tienes acceso a datos etiquetados para el entrenamiento supervisado, o planeas utilizar técnicas de aprendizaje no supervisado?

Respuesta: Al igual que el LLM, el algoritmo de aprendizaje de los modelos es una característica propia y crucial de las herramientas o soluciones que se busca adquirir e implementar. La elección de los modelos y su aprendizaje recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos. No obstante, como se especifica en las actividades 4 y 5, las herramientas y/o soluciones serán probadas y ajustadas, incluyendo allí la posibilidad de variar los algoritmos de aprendizaje con miras a mejoras en el rendimiento en el contexto específico de la VT

8. ¿Qué métricas de evaluación se utilizarán para medir el rendimiento del modelo? ¿Exactitud, precisión, recall, F1-score, entre otras?

Respuesta: No existen métricas de evaluación de rendimientos de Machine Learning o de Inteligencia Artificial adoptadas en la Entidad.

9. ¿Cómo manejarás los desafíos específicos de cada campo (geología, geofísica, geoespacial) en términos de terminología y lenguaje técnico? ¿Será necesario incorporar conocimiento experto en el proceso de clasificación?

Respuesta: Según lo indicado en el apartado "PERSONAL MINIMO", es necesario la incorporación de geólogos en el personal en la ejecución del proyecto. Lo anterior debido a la naturaleza de la información, la cual en su mayoría corresponde a información derivada de contratos, convenios y proyectos desarrollados por la VT.

10. ¿Cómo se integrará el modelo en un sistema o aplicación existente? ¿Se proporcionará una interfaz de usuario para interactuar con los resultados de la clasificación?

Respuesta: La interfaz de usuario (IU) es una propiedad fundamental de las herramientas o soluciones que se busca adquirir e implementar. Es responsabilidad de los interesados proveer información detallada sobre la IU en la propuesta de solución. Este mismo será el componente de interacción directa entre los usuarios y la herramienta que busca dar solución a las necesidades de la organización. Al igual que los modelos empleados, las IU serán probadas con usuarios reales para evaluar la facilidad de uso, diseño, accesibilidad, funcionalidad y eficiencia. Finalmente, se realizarán ajustes a partir de las pruebas de usabilidad, si es necesario. Es importante mencionar que la interfaz de usuario debe alinearse al manual de imagen de la Entidad o a los lineamientos que establezca la dependencia encargada de las comunicaciones internas. En el momento no hay otros lineamientos de experiencia de usuario UX.

11. ¿Cómo se actualizará y mantendrá el modelo con el tiempo? ¿Se necesita un sistema de retroalimentación para mejorar la precisión y relevancia de la clasificación?

Respuesta: Tal como se indica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”, la solución tecnológica debe incluir servicios de mantenimiento y soporte. Los interesados deben detallar los servicios que se brindarán junto con la herramienta o solución, incluyendo allí los de soporte y actualización de los modelos.

12. ¿Cuáles son los requisitos de seguridad y privacidad de los datos que se clasificarán? ¿Se necesita anonimización o cifrado de información sensible?

Respuesta: Los interesados deben incluir en su propuesta lo relacionado a el control de acceso (roles, responsabilidades, mecanismos de control de acceso, monitoreo de accesos), protección de datos (medidas contra intrusiones, robo, pérdida o daño, uso indebido, copias de seguridad y planes de recuperación ante desastres), capacitación frente a políticas de seguridad y privacidad de datos relacionados a la herramienta/solución a implementar, plan de respuesta a incidentes, revisión y actualización de plan de gobernanza de datos. En el caso en que la implementación de la herramienta y/o solución involucre flujo o tránsito de datos e información entre o hacia sistemas externos a la organización, se requiere especificar los detalles de encriptación, anonimización y/o seudonimización que garanticen la gobernanza de los datos.

Preguntas para la indexación documental

1. ¿Cuál es el alcance de la información documental que se va a indexar? ¿Incluye informes técnicos, investigaciones científicas, políticas gubernamentales, datos de energías renovables, etc.?

Respuesta: Tal como se especifica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” y en la Actividad 2, la información a indexar corresponde a tipo documental la cual no requiere software especializado o petrotécnico para su visualización y manejo. En particular, esta información tiene contenido técnico y no técnico que encapsula los aspectos más destacados y cruciales de los contratos, convenios y proyectos gestionados por la VT.

“Dicha información puede corresponder a imágenes de campo (afloramientos rocosos, muestras de roca, núcleos, ripios, fotografías de secciones delgadas, puntos de agua, entre otros), información derivada de perforaciones y pozos (curvas de registros, ensayos y pruebas de pozos, geoquímica de rocas, hidrogeoquímica, análisis isotópicos, entre otros), perfiles e imágenes sísmicas, imágenes espectrales e hiperespectrales, imágenes de tomografía, DRX, FRX, SEM, espectroscopía Raman, secciones estructurales, mapas estructurales, columnas estratigráficas, bloques diagrama, batimetrías y geofísica marina, cartografía (geológica, geomorfológica, geofísica, temática, entre otras), datos de sensores remotos e imágenes satelitales, entre otros.”

“Es fundamental que durante la implementación se reconozca tanto la información documental técnica como la no técnica que pueda encapsular los aspectos más destacados cruciales de los contratos, convenios y proyectos gestionados por la VT. Esto abarca desde estudios previos, sondeos de mercado, comunicaciones internas, minutas, actas de inicio, actas de finalización, actas de liquidación, otrosí, actas de entregas, hasta productos en formatos no técnicos como informes finales, informes de actividades y supervisión. No obstante, es crucial señalar que, dada la disposición actual de la información, la cual mayormente carece de estructura, se torna imperativo identificar y ubicar la información

definitiva y esencial almacenada en nuestro directorio colaborativo. Esto se debe a la posibilidad de que existan versiones preliminares de cada uno de los documentos mencionados anteriormente.”

2. ¿Cuáles son los criterios de búsqueda más importantes para los usuarios? ¿Palabras clave, fechas, tipos de documentos, temas específicos relacionados con la transición energética?

Respuesta: Esta pregunta se encuentra interconectada con las preguntas 3 y 5 del apartado “Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos”.

Ahora bien, las principales búsquedas que se desarrollan en la VT corresponden a objetos, alcance, fechas, palabras claves temáticas, áreas de estudio, ubicación geográfica, actas o documentos precontractuales-contractuales-postcontractuales, temáticas abordadas, metodología, conclusiones, productos, entre otros aspectos técnicos, de los contratos, convenios y proyectos desarrollados por la VT. De igual forma, como se especifica en las actividades 4 y 5, las herramientas y/o soluciones serán probadas y ajustadas, incluyendo búsquedas específicas por usuarios funcionales finales, para ajustar o incorporar nuevos criterios de búsqueda, metadatos, características y/o etiquetas en los algoritmos o modelos empleados para mejorar el rendimiento de las herramientas y/o soluciones.

3. ¿Qué tecnologías de indexación y búsqueda se consideran para implementar la solución? ¿Motor de búsqueda convencional, sistemas de recuperación de información basados en inteligencia artificial, tecnologías de procesamiento del lenguaje natural, etc.?

Respuesta: La elección de las tecnologías de indexación y búsqueda de información recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos.

4. ¿Cómo se adquirirán los documentos para su indexación? ¿Se utilizarán rastreadores web, sistemas de gestión de documentos, repositorios digitales, etc.?

Respuesta: Esta pregunta se encuentra interconectada a la pregunta 4 del apartado “Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos”

5. ¿Qué métodos se utilizarán para procesar y normalizar los documentos antes de la indexación? ¿Extracción de texto, eliminación de metadatos irrelevantes, detección de idiomas, etc.?

Respuesta: Al igual que se indicó en la pregunta 7 del apartado “Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos”, la elección de los modelos y su aprendizaje, el preprocesamiento de la información y la elección de los metadatos a extraer durante la indexación de la información, recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos.

6. ¿Cómo se estructurará y organizará la información indexada? ¿Por categorías, etiquetas, fechas, o algún otro criterio?

Respuesta: La información indexada será el insumo principal para las búsquedas de información dentro del proyecto. La forma en la que esta información se gestionará y accederá por el buscador hará parte de la propuesta desarrollada por los interesados.

Sin embargo, como se indica en el sondeo, los metadatos resultantes de la indexación son uno de los productos esperados del proyecto. Por lo tanto, los interesados también deberán proponer cómo se facilitarán estos metadatos.

7. ¿Qué nivel de granularidad se aplicará en la indexación? ¿Se indexarán documentos completos o partes específicas dentro de los documentos?

Respuesta: Según la necesidad planteada en la organización, la granularidad en la indexación que se espera corresponde a la más detallada posible, que capture la mayor cantidad de información posible para que las búsquedas, acceso y consultas sean más eficientes.

8. ¿Qué métricas se utilizarán para evaluar la eficiencia de la búsqueda? ¿Tiempo de respuesta, precisión, cobertura, relevancia de los resultados, etc.?

Respuesta: Las métricas para evaluar la eficiencia de la herramienta y/o solución corresponden a objetivos y criterios claros y medibles de éxito. Estos como mínimo deben evaluar la efectividad del indexador y buscador en términos de eficiencia de almacenamiento (tamaño total de los datos almacenados versus espacio físico disponible), rendimiento del acceso a los datos (velocidad y capacidad de respuesta al realizar consultas, recuperación y carga de información asociada al resultado de la búsqueda), tasa de transferencia de datos, capacidad de procesamiento adicional (indexación), capacidad de manejo de cargas de trabajo variables y simultáneas (indexación y búsqueda), capacidad de recepción de nuevos usuarios e información, disponibilidad y fiabilidad de los resultados de consultas, capacidad de mantener la integridad y consistencia de datos, tasa de error de datos, falla de software y hardware, auditoría de seguridad, cumplimiento de regulaciones de privacidad y protección de datos, costo total de propiedad del indexador y buscador (incluyendo adquisición, implementación, mantenimiento y operación del sistema), entre otros. Sin embargo, los interesados pueden proponer nuevas métricas, así como considerar que algunas de las mencionadas anteriormente no son representativas.

Es importante resaltar que no existen métricas de evaluación de eficiencia de software adoptadas en la Entidad, no obstante, se requiere que las aplicaciones que se implementen tengan en cuenta las consideraciones del modelo de calidad de la norma ISO 25000

9. ¿Cómo se asegurará la seguridad y privacidad de la información indexada? ¿Se aplicarán medidas de cifrado, control de acceso, anonimización, etc.

Respuesta: Esta pregunta se encuentra interconectada a la pregunta 11 del apartado "Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos"

10. ¿Cómo se manejará la escalabilidad y mantenimiento del sistema a medida que la cantidad de documentos crezca? ¿Se implementarán técnicas de distribución y paralelización?

Respuesta: Según lo indicado en el apartado "ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR", el software/herramienta/aplicación de indexación y búsqueda debe garantizar la posibilidad de personalización y escalabilidad, bien sea a mayores volúmenes de datos o distintos tipos de información. No obstante, la elección de las técnicas para garantizar estas capacidades corresponde a una responsabilidad de los interesados y debe ser incluida en la propuesta de solución.

11. ¿Se considerará la expansión futura de la solución para incluir más tipos de documentos o fuentes de información? ¿Se planificará una arquitectura flexible y modular?

Respuesta: Según lo indicado en la actividad 3 del apartado "ACTIVIDADES", el software/herramienta/aplicación debe ser capaz de escalar o expandirse, la manera en que estas capacidades sean desarrolladas será elección de los interesados. No obstante, deben ser sustentadas por diagramas de deslignes y arquitecturas de despliegues, al igual que otros documentos relevantes.

Preguntas para la experiencia

1. ¿La experiencia de integrantes del equipo de trabajo en el sector Oil & Gas es válida para el requisito de la validación de la experiencia?

Respuesta: Se recomienda la participación de al menos un profesional en geología o ingeniería geológica debido a la naturaleza técnica de los datos e información relacionados con el proyecto. Para los demás perfiles del personal mínimo, se recomienda consultar la tabla "PERSONAL MÍNIMO" que se encuentra disponible en el sondeo de mercado. La tabla detalla la experiencia específica y las habilidades requeridas para cada perfil.

PREGUNTAS PRESENTADAS POR: GEO OIL ENERGY S.A.S

1. En el documento "1572728 SM para la Indexación de la Información TE", bajo el ítem de "Aspectos técnicos y actividades a ejecutar", específicamente en "Creación de metadatos", solicitamos amablemente a la ANH aclarar a que se refieren con "tipo y versión de software del cual se deriva información contenida en los documentos".

Respuesta: Si es posible, identificar el software técnico específico y su versión utilizados en la creación de imágenes, gráficos o información técnica contenida en la documentación indexada. Esta práctica facilitará una búsqueda eficiente de información, dado que hay productos derivados de ciertos softwares que se pueden vincular a temas o etiquetas específicas. Además, asociar

un metadato a esta información puede orientar a los usuarios finales que deseen acceder a archivos gestionados en software petrotécnicos.

2. Solicitamos a la ANH aclarar si, la información a utilizar ya se tiene lista en una sola ubicación o, si por el contrario, dentro del desarrollo del servicio es necesario realizar una búsqueda general de la información dentro de los servidores de la ANH para seleccionar la información pertinente a indexar

Respuesta: La información actualmente está almacenada en un servidor físico, al cual se accede a través de una ubicación de red que corresponde a una carpeta compartida. En su mayoría, esta información no está estructurada, lo que conlleva a que se requiera un preprocesamiento para identificar, clasificar y validar la veracidad, integridad, confiabilidad y consistencia de la información. Solo así se podrá recopilar la información correcta para la implementación de la herramienta de indexación y búsqueda. Es fundamental que los interesados propongan maneras eficientes e innovadoras para llevar a cabo este preprocesamiento.

3. Considerando que el servicio es para indexación de la industria del O&G, solicitamos a la ANH añadir a los perfiles de Director del proyecto y experto y analista de bases de datos la profesión de geólogo o ingeniero geológico. Esto dado que en la industria se cuenta con profesionales de ambas carreras que durante el desarrollo de su profesión han trabajado en proyectos de manejo de información e indexación de información de la industria del O&G, cuya experiencia combinada con el conocimiento técnico de los datos a indexar, serían de mucho valor en el desarrollo del proyecto.

Respuesta: En caso que se desarrolle un proceso de selección y una posterior contratación, esta observación se tendrá en cuenta para la construcción de los términos de los mismos.

4. Dentro de los productos a entregar, la ANH menciona. “Suministro del software/herramienta/aplicación, con el debido licenciamiento y derechos de uso para la VT, que pueda ser usada con la infraestructura tecnológica actual, cuente con soporte y sea puesta en funcionamiento”. Solicitamos a la ANH aclarar el tiempo de licenciamiento y el tiempo de soporte que esperan adquirir del software/herramienta/aplicación.

Respuesta: Tal como se indica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”, la solución tecnológica debe incluir servicios de mantenimiento y soporte. Los interesados deben detallar los servicios que se brindarán junto con la herramienta o solución, incluyendo allí los de soporte y actualización de los modelos. No obstante, se sugiere que el licenciamiento sea perpetuo con una vigencia de soporte y mantenimiento de mínimo un año.

5. Solicitamos a la ANH aclarar si el software/herramienta/aplicación puede ser una solución en la nube o, si, por el contrario, el despliegue tiene que ser completamente en los servidores de la ANH.

Respuesta: Se sugiere que la solución que se adquiera sea On-premise y Standalone, de tal forma que se pueda utilizar en otros procesos y no tenga una limitación por tiempo de servicio.

6. Dado que las tecnologías actuales permiten el trabajo remoto colaborativo de una manera eficiente, solicitamos a la ANH que modifique el lugar de ejecución del proyecto y permita el trabajo remoto para las labores de ingesta de la información. Con la salvedad que, para la instalación de la herramienta, pruebas de funcionamiento y presentación de resultados, miembros del equipo se desplacen a las instalaciones de la ANH para tener las reuniones pertinentes.

Respuesta: En caso de que se desarrolle un proceso de selección y una posterior contratación, esta observación se tendrá en cuenta para la construcción de los términos de los mismos. Sin embargo, el sondeo de mercado solo especifica el personal mínimo requerido y el lugar de ejecución del proyecto, sin contemplar la asistencia permanente durante su desarrollo. Es importante destacar que, como se menciona en el enunciado de la pregunta, pueden surgir momentos críticos durante la ejecución del proyecto que requieran la presencia de cierto personal del proveedor de la solución.

PREGUNTAS PRESENTADAS POR: GLOBANT

1. ¿Qué tipos de documentos y formatos se deben indexar (PDF, Word, imágenes, etc.)?

Respuesta: Tal como se especifica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” y en la Actividad 2, la información a indexar corresponde a tipo documental la cual no requiere software especializado o petrotécnico para su visualización y manejo, esta puede incluir archivos de imagen, archivos PDF, documentos escaneados, documentos de texto, Word, Excel, PowerPoint, entre otros. En particular, esta información tiene contenido técnico y no técnico que encapsula los aspectos más destacados y cruciales de los contratos, convenios y proyectos gestionados por la VT.

2. ¿Cuál es el volumen estimado de documentos a procesar?

Respuesta: Tal como se indica en el apartado “RECURSOS TECNOLÓGICOS”, los interesados deben presentar tres escenarios y/o propuestas de soluciones basados en el procesamiento de los siguientes volúmenes de información documental: 5TB, 10TB, 20TB.

3. ¿Requieren que todos los documentos estén en un único repositorio o se puede usar una estrategia de particionar por temas?

Respuesta: Se sugiere que la información se encuentre almacenada en un único sistema de almacenamiento centralizado. No obstante, los interesados pueden proponer distintas estrategias innovadoras de almacenamiento de la información.

4. Además de los documentos, ¿existen datos estructurados que deban incorporarse para indexar? (por ejemplo, bases de datos, archivos Excel, etc)

Respuesta: Tal como se especifica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” y en la Actividad 2, la información a indexar se encuentra en su mayoría no estructurada y corresponde documentos que no requieren software especializado o petrotécnico

para su visualización y manejo, como un sistema de gestión de bases de datos (DBMS). Ahora bien, esta información puede incluir archivos de imagen, archivos PDF, documentos escaneados, documentos de texto, Word, Excel, PowerPoint, entre otros. En particular, esta información tiene contenido técnico y no técnico que encapsula los aspectos más destacados y cruciales de los contratos, convenios y proyectos gestionados por la VT.

5. ¿Existen documentos en idiomas distintos al español que requieran ser indexados?

Respuesta: Sí, disponemos de información en varios idiomas. Sin embargo, se espera que el software/herramienta/aplicación debe ser capaz de procesar información y realizar consultas que incluyan, como mínimo, palabras y expresiones en inglés. No obstante, se sugiere a los interesados reconocer y familiarizarse con las principales palabras o expresiones en otros idiomas comúnmente usadas en el contexto técnico que rodea la información del proyecto.

6. ¿Donde residen estos documentos hoy en día? Por ejemplo, GDrive, Sharepoint, Folders en una red interna, algún otro sistema documental que estén usando en la actualidad

Respuesta: La información actualmente está almacenada en un servidor físico, al cual se accede a través de una ubicación de red que corresponde a una carpeta compartida.

7. ¿Hay un esquema de metadatos predefinido o estándares específicos que se deben seguir?
¿Qué atributos específicos de metadatos son prioritarios para la organización?

Respuesta: No existen esquemas de metadatos en la Entidad. Ahora bien, la manera en que se lleva a cabo la tipología, clasificación y catalogación documental, lo cual es fundamental en la indexación y búsqueda de información, no obedece a una manera en particular de hacerlo, ni a unas etiquetas, metadatos y categorías específicas. La elección del esquema, estándares y los metadatos a extraer recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos. Igualmente, la propuesta de los interesados en materia de lo anteriormente mencionado debe considerar que los metadatos resultantes de la indexación son uno de los productos esperados del proyecto. Por lo tanto, los interesados también deberán proponer cómo se facilitarán estos metadatos.

8. ¿Qué sistemas o plataformas actuales se deben considerar para la integración?

Respuesta: La solución/aplicación debe tener la capacidad de poder integrarse con la plataforma tecnológica de la ANH que cuenta con las siguientes características específicas:

- ✓ Sistemas Operativos de Microsoft (Windows Server 2019) y Linux (Ubuntu server o de libre distribución)
- ✓ Sistema Manejador de Base de Datos: Microsoft SQL Server 2016
- ✓ IaaS - PaaS con Servicios de Microsoft Azure, no obstante, se sugiere que la solución sea On-premise.

9. ¿Se requiere compatibilidad con algún estándar específico de API (REST, SOAP, GraphQL, etc.)?

Respuesta: La Arquitectura de referencia de desarrollo define que los servicios estén basados en protocolos y estándares (REST, JSON, HTTP, SSL) para intercambiar información entre aplicaciones.

10. ¿Qué nivel de complejidad se espera en las búsquedas semánticas?

Respuesta: Las principales búsquedas que se desarrollan en la VT corresponden a objetos, alcance, fechas, palabras claves temáticas, áreas de estudio, ubicación geográfica, actas o documentos precontractuales-contractuales-postcontractuales, temáticas abordadas, metodología, conclusiones, productos, entre otros aspectos técnicos, de los contratos, convenios y proyectos desarrollados por la VT. Estas consultas estarán en su mayoría enmarcadas en contextos geológicos, por lo cual es necesario enfocar las búsquedas de manera que se entiendan relaciones de palabras técnicas relacionadas a los temas anteriormente mencionados.

11. ¿Se requieren capacidades multilingües para el procesamiento de lenguaje natural?

Respuesta: Tal como se indica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” del sondeo de mercado, los modelos de NLP del software/herramienta/aplicación deben ser capaces de procesar información y consultas que incluyan, como mínimo, palabras y expresiones en inglés. No obstante, se sugiere a los interesados reconocer y familiarizarse con las principales palabras o expresiones en otros idiomas que se utilizan habitualmente en el contexto técnico que rodea la información del proyecto.

12. ¿Cómo se debe tratar la documentación repetida (eliminar, archivar, marcar, etc.)?

Respuesta: Se recomienda eliminar la documentación duplicada del conjunto de datos proporcionado para el proyecto antes de indexar la información. Esto permitirá mejorar la eficiencia del proceso de indexación y evitará que la información duplicada consuma recursos y sea incluida en las distintas etapas del proyecto.

Los duplicados deben ser referenciados y consolidados para que la Entidad pueda extraerlos de sus repositorios. Esto puede ser realizado creando una lista de todos los documentos duplicados y sus ubicaciones.

13. ¿Qué tipo de insights son de interés para la organización?

Respuesta: Tal como se indica en la descripción de la necesidad, se requiere una solución tecnológica que permita búsquedas, acceso y consultas eficientes de información. Ahora bien, la manera en que se lleva a cabo la tipología, clasificación y catalogación documental, lo cual es fundamental en la indexación y búsqueda de información, no obedece a una manera en particular de hacerlo, ni a unas etiquetas, insights y categorías específicas. Sin embargo, la propuesta de los interesados en materia de lo anteriormente mencionado debe responder eficientemente a la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos. Para orientar la elección de estos insights por parte de los interesado, las principales búsquedas de información en la VT incluyen objetos, alcance, fechas, palabras claves temáticas, áreas de

estudio, ubicación geográfica, actas o documentos precontractuales-contractuales-postcontractuales, temáticas abordadas, metodología, conclusiones, productos, entre otros aspectos técnicos, de los contratos, convenios y proyectos desarrollados por la VT.

14. ¿Existen preguntas o problemas específicos que la herramienta de análisis debería resolver?

Respuesta: El propósito principal de la indexación de la información es la búsqueda, acceso y consulta eficiente de información, esto mediante una herramienta de indexación y búsqueda que integre nuevas tecnologías de la información y ciencia de la computación. Ahora bien, las principales búsquedas que se desarrollan en la VT corresponden a objetos, alcance, fechas, palabras claves temáticas, áreas de estudio, ubicación geográfica, actas o documentos precontractuales-contractuales-postcontractuales, temáticas abordadas, metodología, conclusiones, productos, entre otros aspectos técnicos, de los contratos, convenios y proyectos desarrollados por la VT. Estas consultas estarán en su mayoría enmarcadas en contextos geológicos, por lo cual es necesario enfocar las búsquedas de manera que se entiendan preguntas con relaciones de palabras técnicas relacionadas a los temas anteriormente mencionados.

15. ¿Qué nivel de precisión se requiere para el OCR?

Respuesta: El OCR es una capacidad fundamental de las herramientas o soluciones que se busca adquirir e implementar. Ahora bien, la precisión puede variar dependiendo del tipo de documento, calidad, fuente, complejidad del diseño, idioma del texto y la naturaleza de la información que sea requerida mediante una consulta. En especial, la precisión requerida para el OCR y su tolerancia dependerá de los factores anteriormente mencionados, no obstante, si la naturaleza o importancia de la información consultada corresponde específicamente a datos críticos o legalmente vinculantes de carácter precontractual-contractual-postcontractual (valores, entidades, objetos, productos, fechas, alcance, áreas de estudio, entre otros), la precisión requerida será alta (95%-100%), esto con el fin de evitar resultados de búsquedas erróneos que puedan tener consecuencias negativas para la Entidad. En el caso que la precisión del OCR sea media o baja, debido a los factores anteriormente mencionados, u otros, se sugiere que sea advertido el usuario de la herramienta o solución del nivel de precisión. La elección de las estrategias para mejorar la precisión del OCR recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos.

16. ¿Se necesitan capacidades de reconocimiento de imágenes o análisis de video?

Respuesta: Según lo indicado en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”, el software/herramienta/aplicación de indexación y búsqueda debe ser capaz de indexar información a partir de diferentes tipos de imágenes de información técnica geológica, geofísica o geoespacial. Pueden corresponder a imágenes de campo en diferentes escalas de detalle, imágenes de muestras de campo, imágenes de secciones delgadas, imágenes sísmicas, imágenes estructurales, entre otros.

“VISIÓN COMPUTACIONAL: Reconocimiento, extracción y categorización de información técnica de las 3G (Geología, Geofísica, Geoespacial) mediante reconocimiento y análisis de

imágenes. Dicha información puede corresponder a imágenes de campo (afloramientos rocosos, muestras de roca, núcleos, ripios, fotografías de secciones delgadas, puntos de agua, entre otros), información derivada de perforaciones y pozos (curvas de registros, ensayos y pruebas de pozos, geoquímica de rocas, hidrogeoquímica, análisis isotópicos, entre otros), perfiles e imágenes sísmicas, imágenes espectrales e hiperespectrales, imágenes de tomografía, DRX, FRX, SEM, espectroscopía Raman, secciones estructurales, mapas estructurales, columnas estratigráficas, bloques diagrama, batimetrías y geofísica marina, cartografía (geológica, geomorfológica, geofísica, temática, entre otras), datos de sensores remotos e imágenes satelitales, entre otros.”

17. ¿Para la extracción de información técnica a partir de imágenes están previendo el entrenamiento de modelos de computer vision basados en las imágenes que cuentan actualmente?

Respuesta: La solución preferiblemente debe contar con algoritmos y modelos entrenados y especializados en la extracción de información desde conjuntos de datos técnicos como los indicados en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”. De no contar con estos, los mismos deben ser entrenados y puestos en producción dentro de los tiempos previstos para la ejecución del contrato.

18. ¿Qué volumen de imágenes se cuenta en caso de tener que entrenar un modelo para reconocer la información que se quieren obtener de las mismas?

Respuesta: No hay un volumen de imágenes estimado, dado que las mismas están inmersas dentro de los volúmenes de información documental. Se debe realizar una propuesta económica atendiendo a los volúmenes de información documental previstos: 5TB, 10TB y 20TB.

19. ¿Cuál es el nivel de soporte técnico requerido (24/7, horario laboral, etc.)?

Respuesta: El soporte técnico deberá corresponder al horario laboral de la ANH, de 7:00 – 17:00 horas (GMT-5).

20. ¿Se espera un acuerdo de nivel de servicio (SLA) específico?

Respuesta: En caso de realizarse el proyecto se espera un acuerdo de nivel de servicio (SLA) cuyos términos serían definidos en etapas posteriores.

21. ¿Hasta qué punto se espera que la solución sea personalizable?

Respuesta: Según lo indicado en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”, el software/herramienta/aplicación de indexación y búsqueda debe garantizar la posibilidad de personalización y escalabilidad, bien sea a mayores volúmenes de datos o distintos tipos de información y documentos (formatos como shp, mxd, entre otros). No obstante, la elección de las técnicas para garantizar estas capacidades corresponde a una responsabilidad de los interesados y deben ser incluidas en la propuesta de solución. Adicionalmente, deben ser sustentadas por diagramas de deslignes y arquitecturas de despliegues, al igual que otros documentos relevantes.

22. ¿Se anticipa un crecimiento en el volumen de datos o en el número de usuarios que podría afectar la escalabilidad?

Respuesta: Según lo indicado en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”, el software/herramienta/aplicación de indexación y búsqueda debe garantizar la posibilidad de personalización y escalabilidad, bien sea a mayores volúmenes de datos o distintos tipos de información y documentos (formatos como shp, mxd, entre otros). No obstante, la elección de las técnicas para garantizar estas capacidades corresponde a una responsabilidad de los interesados y deben ser incluidas en la propuesta de solución. Adicionalmente, deben ser sustentadas por diagramas de desligues y arquitecturas de despliegues, al igual que otros documentos relevantes.

23. ¿Qué funcionalidades específicas se esperan del chatbot o de las modalidades de búsqueda inmersiva?

Respuesta: Las funciones específicas del chatbot deben atender la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos elegidos por los interesado para las herramientas y/o soluciones. En términos generales, el chatbot debe ser capaz de procesar y generar de lenguaje natural relacionado con las principales búsquedas en la VT. Estas búsquedas incluyen:

- Objetos
- Alcance
- Fechas
- Palabras claves temáticas
- Áreas de estudio
- Ubicación geográfica
- Actas o documentos precontractuales-contractuales-postcontractuales
- Temáticas abordadas
- Metodología
- Conclusiones
- Productos

Entre otros aspectos técnicos de los contratos, convenios y proyectos desarrollados por la VT. Adicionalmente, debe ser capaz de rastrear la fuente de la información información, si así lo solicita el usuario. Esto permitirá a los usuarios verificar la información proporcionada por el chatbot y acceder a información adicional si lo necesitan.

Por ejemplo:

Usuario: ¿Qué contratos desarrollados por la VT han hecho post mortem en el área del VMM?

Chatbot: Los contratos X, XX y XXX reportan actividades de postmortem en su metodología y resultados.

Usuario: ¿En qué año se firmó el contrato X?

Chatbot: El contrato con la empresa X se firmó en el año 2023. La información se encuentra disponible en el acta de la reunión del 14 de marzo de 2023, página 5.

Usuario: ¿Puedo ver el acta de la reunión?

Chatbot: Sí, puede acceder al acta de la reunión haciendo clic en el siguiente enlace: [enlace al acta de la reunión]

24. ¿Hay preferencias en cuanto a la interfaz de usuario o la experiencia del usuario (UX)?

Respuesta: La interfaz de usuario debe alinearse al manual de imagen de la Entidad o a los lineamientos que establezca la dependencia encargada de las comunicaciones internas. En el momento no hay otros lineamientos de experiencia de usuario UX.

25. ¿La interfaz de usuario deberá ser web responsive o además se requieren aplicaciones móviles nativas para iOS y Android?

Respuesta: No se requiere desarrollo de aplicaciones móviles. En cuanto a la interfaz de usuario se busca que pueda ser accesible desde una aplicación de escritorio nativa.

26. ¿Existen requisitos específicos de seguridad de la información o cumplimiento de normativas (como GDPR, HIPAA, etc.)?

Respuesta: Los requisitos en materia de seguridad de la información corresponden al cumplimiento de las políticas de seguridad y privacidad de la información de la Entidad, las cuales, se alinean al Modelo de Seguridad y Privacidad de la Información - MSPI emitido por el Ministerio de Tecnologías y las Comunicaciones – MINTIC, quien regula a las entidades de la rama ejecutiva en cuando a lineamientos tecnológicos. Cabe anotar que el MSPI se encuentra alineado a su vez con las buenas prácticas ISO 27001.

Adicionalmente, se recomienda revisar la pregunta 11 del apartado “Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos” del presente documento.

27. ¿Qué requerimientos tienen en cuanto a la privacidad y confidencialidad de los datos?

Respuesta: En Concordancia con lo mencionado anteriormente, los requisitos corresponden al cumplimiento de las políticas de seguridad y privacidad de la información de la Entidad, las cuales, se alinean al Modelo de Seguridad y Privacidad de la Información - MSPI emitido por el Ministerio de Tecnologías y las Comunicaciones – MINTIC quien regula a las entidades de la rama ejecutiva en cuando a lineamientos tecnológicos. En cuanto a la confidencialidad de los datos, se debe realizar la tipología, clasificación y catalogación documental de los activos de información del proceso respectivo. Esto permitirá determinar la confidencialidad que se requiere.

Adicionalmente, se recomienda revisar la pregunta 11 del apartado “Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos” del presente documento.

28. ¿Cuál es el plazo de ejecución del proyecto?

Respuesta: El plazo de ejecución se ha establecido hasta un máximo de siete (7) meses, sin sobrepasar el 16 de diciembre de 2024. Sin embargo, se trata de un tiempo estimado teniendo en cuenta la vigencia presupuestal. En caso de realizarse el proyecto este plazo puede estar sujeto a modificaciones.

29. ¿Hay un presupuesto estimado o un rango de costos esperado para la solución?

Respuesta: El presente sondeo de mercado tiene como fin establecer un presupuesto de acuerdo con un análisis económico de las diferentes propuestas de los interesados y el cubrimiento de los requerimientos técnicos específicos de la VT y las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado.

30. ¿Cuáles son los criterios de evaluación y ponderación para seleccionar la propuesta ganadora?

Respuesta: En caso que se ejecute el proyecto, el mecanismo de realización tendrá criterios de evaluación y ponderación específicos, según del tipo de proceso de selección y contratación en el que se enmarque.

31. ¿Se requieren pruebas piloto o fases de validación antes de la implementación completa?

Respuesta: Tal como se especifica en las actividades 4 y 5, las herramientas y/o soluciones serán probadas y ajustadas, incluyendo allí la posibilidad de validar su rendimiento con miras a lograr una implementación exitosa en el contexto específico de las necesidades de la VT a las que se busca dar solución.

32. ¿Se espera que la propuesta incluya un plan de capacitación para los usuarios?

Respuesta: Tal como se indica en la actividad 6, los interesados deberán contemplar en sus propuestas la realización de capacitaciones a usuarios funcionales finales de la Entidad. Se sugiere revisar los entregables asociados a esta actividad que se especifican en el sondeo de mercado.

33. ¿Cómo se deben manejar los datos sensibles o confidenciales durante la indexación y búsqueda?

Respuesta: De conformidad con lo enunciado en puntos anteriores y teniendo en cuenta la Ley 1581 de 2012 y sus reglamentaciones, se debe dar cumplimiento al tratamiento de datos conforme la política de tratamiento de datos personales de la Entidad y la normativa vigente. En cuanto a la confidencialidad de los datos, se debe realizar la tipología, clasificación y catalogación

documental de los activos de información del proceso respectivo. Esto permitirá determinar la confidencialidad que se requiere.

Adicionalmente, se recomienda revisar la pregunta 11 del apartado “Preguntas para la construcción de un modelo ML o LLM que nos permita la clasificación de documentos” del presente documento.

34. ¿Existen políticas de retención de datos que deban cumplirse?

Respuesta: La retención de información para la documentación de la Entidad aplica conforme los lineamientos del Archivo General de la Nación y las Tablas de Retención Documental así como la necesidad de históricos de la dependencia. En caso de realizarse el proyecto estos lineamientos pueden estar sujetos a modificaciones.

35. ¿La solución debe ser alojada en la nube, on-premise o híbrida?

Respuesta: Se sugiere que la solución que se adquiera sea On-premise y Standalone, de tal forma que se pueda utilizar en otros procesos y no tenga una limitación por tiempo de servicio.

36. ¿Hay preferencias o restricciones tecnológicas (por ejemplo, Cloud Service Providers, sistemas operativos, bases de datos, etc.)?

Respuesta: La plataforma tecnológica de la ANH tiene las siguientes características:

- ✓ Sistemas Operativos de Microsoft (Windows Server 2019) y Linux (Ubuntu server o de libre distribución)
- ✓ Sistema Manejador de Base de Datos: Microsoft SQL Server 2016
- ✓ IaaS - PaaS con Servicios de Microsoft Azure, no obstante, se sugiere que la solución sea On-premise.

37. ¿Qué estrategias de respaldo y recuperación de datos se deben implementar?

Respuesta: Las estrategias de respaldo y de recuperación deben ser propuestas y recomendadas por el implementador de la solución en los documentos de aseguramiento de aplicaciones

38. ¿Cuáles son los criterios específicos de aceptación para la solución entregada?

Respuesta: Las métricas para evaluar la eficiencia y los criterios de aceptación de la herramienta y/o solución corresponden a objetivos y criterios claros y medibles de éxito. Estos como mínimo deben evaluar la efectividad del indexador y buscador en términos de eficiencia de almacenamiento (tamaño total de los datos almacenados versus espacio físico disponible), rendimiento del acceso a los datos (velocidad y capacidad de respuesta al realizar consultas, recuperación y carga de información asociada al resultado de la búsqueda), tasa de transferencia de datos, capacidad de procesamiento adicional (indexación), capacidad de manejo de cargas

de trabajo variables y simultaneas (indexación y búsqueda), capacidad de recepción de nuevos usuarios e información, disponibilidad y fiabilidad de los resultados de consultas, capacidad de mantener la integridad y consistencia de datos, tasa de error de datos, falla de software y hardware, auditoria de seguridad, cumplimiento de regulaciones de privacidad y protección de datos, costo total de propiedad del indexador y buscador (incluyendo adquisición, implementación, mantenimiento y operación del sistema), entre otros. Sin embargo, los interesados pueden proponer nuevas métricas o criterios de aceptación, así como considerar que algunas de las mencionadas anteriormente no son representativas.

Es importante resaltar que no existen métricas de evaluación de eficiencia de software adoptadas en la Entidad, no obstante, se requiere que las aplicaciones que se implementen tengan en cuenta las consideraciones del modelo de calidad de la norma ISO 25000

39. ¿Es esencial que la inteligencia artificial de búsqueda esté dotada de la habilidad para rastrear información que sea contextualmente relevante dentro de los documentos, interpretando así el sentido y el contexto del contenido, en vez de restringirse únicamente a búsquedas indexadas estándar?

Respuesta: Tal como se especifica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR”, el software/herramienta/aplicación debe ser capaz de realizar búsquedas semánticas para lo cual es esencial interpretar el sentido y contexto de la indexación y búsquedas a realizar.

PREGUNTAS PRESENTADAS POR: SERVINFORMACIÓN

1. Creación de metadatos: se solicita a la entidad aclarar si para la creación de metadatos se tiene un estándar (NTC - Dublin Core) o si la entidad tiene definidos los campos a extraer de la información, cuales son obligatorios - opcionales de acuerdo al tipo de información - ¿La ANH tiene definida la forma en que deben ser clasificados y catalogados los documentos en general o el proponente deberá estructurar una?

Respuesta: No existen esquemas de metadatos en la Entidad. Ahora bien, la manera en que se lleva a cabo la tipología, clasificación y catalogación documental, lo cual es fundamental en la indexación y búsqueda de información, no obedece a una manera en particular de hacerlo, ni a unas etiquetas, metadatos y categorías específicas. La propuesta de los interesados en materia de lo anteriormente mencionado debe responder eficientemente a los aspectos técnicos y la necesidad especificados en el sondeo de mercado.

Igualmente, la elección del esquema, estándares y los metadatos a extraer recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos. Los metadatos resultantes de la indexación son uno de los productos esperados del proyecto. Por lo tanto, los interesados también deberán proponer cómo se facilitarán estos metadatos.

2. Tipos de formatos y métricas en los que se encuentra la información.

Respuesta: Tal como se especifica en el apartado “ASPECTOS TÉCNICOS Y ACTIVIDADES A EJECUTAR” y en la Actividad 2 del sondeo de mercado, la información a indexar corresponde a tipo documental la cual no requiere software especializado o petrotécnico para su visualización y manejo, esta puede incluir archivos de imagen, archivos PDF, documentos escaneados, documentos de texto, Word, Excel, PowerPoint, entre otros. En particular, esta información tiene contenido técnico y no técnico que encapsula los aspectos más destacados y cruciales de los contratos, convenios y proyectos gestionados por la VT.

3. Actualmente como se encuentra dispuesta la información a catalogar? está en la nube? on premise?, se tiene información en bases de datos?

Respuesta: La información actualmente está almacenada en un servidor físico, al cual se accede a través de una ubicación de red que corresponde a una carpeta compartida.

4. En cuanto a la búsqueda semántica mediante procesamiento y generación de lenguaje natural (nlp). Solicitamos a la entidad aclarar lo que se espera en cuanto a: “interacción inteligente con la información.”, la interacción comprende el descubrimiento de la información, o que el chatbot responda con una base de conocimiento basada en el archivo a requerir.

Respuesta: Las funciones específicas del chatbot deben atender la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos elegidos por los interesado para las herramientas y/o soluciones. En términos generales, el chatbot debe ser capaz de procesar y generar de lenguaje natural relacionado con las principales búsquedas en la VT. Estas búsquedas incluyen:

- Objetos
- Alcance
- Fechas
- Palabras claves temáticas
- Áreas de estudio
- Ubicación geográfica
- Actas o documentos precontractuales-contractuales-postcontractuales
- Temáticas abordadas
- Metodología
- Conclusiones
- Productos

Entre otros aspectos técnicos de los contratos, convenios y proyectos desarrollados por la VT. Adicionalmente, debe ser capaz de rastrear la fuente de la información información, si así lo solicita el usuario. Esto permitirá a los usuarios verificar la información proporcionada por el chatbot y acceder a información adicional si lo necesitan.

Por ejemplo:

Usuario: ¿Qué contratos desarrollados por la VT han hecho post mortem en el área del VMM?

Chatbot: Los contratos X, XX y XXX reportan actividades de postmortem en su metodología y resultados.

Usuario: ¿En qué año se firmó el contrato X?

Chatbot: El contrato con la empresa X se firmó en el año 2023. La información se encuentra disponible en el acta de la reunión del 14 de marzo de 2023, página 5.

Usuario: ¿Puedo ver el acta de la reunión?

Chatbot: Sí, puede acceder al acta de la reunión haciendo clic en el siguiente enlace: [enlace al acta de la reunión]

Asimismo, los modelos de NLP del software/herramienta/aplicación deben ser capaces de procesar información y consultas que incluyan, como mínimo, palabras y expresiones en inglés. No obstante, se sugiere a los interesados reconocer y familiarizarse con las principales palabras o expresiones en otros idiomas que se utilizan habitualmente en el contexto técnico que rodea la información del proyecto.

5. ¿Cómo se espera el consumo del chatbot, únicamente a través de la plataforma proporcionada, o se contempla la posibilidad de integración con una página web o herramientas de terceros como WhatsApp?

Respuesta: Se sugiere que la interacción usuarios – bot se realice mediante una interfaz de usuario contenida en una aplicación de escritorio nativa que corresponde al software/herramienta/aplicación.

6. En cuanto a la extracción de información - reconocimiento óptico de caracteres, es importante comprender que el tema de la calidad de los escaneos requiere de muy buenos niveles de calidad, para así mismo extraer la información requerida con calidad, por lo que se solicita a la entidad fijar niveles de aceptación para los documentos escaneados. Aunado al ítem anterior es importante definir los idiomas en los que se debe realizar las extracciones y si estos idiomas deben ser tenidos en cuenta en la implementación de la solución de indexación.

Respuesta: El OCR es una capacidad fundamental de las herramientas o soluciones que se busca adquirir e implementar. Ahora bien, la precisión puede variar dependiendo del tipo de documento, calidad, fuente, complejidad del diseño, idioma del texto y la naturaleza de la información que sea requerida mediante una consulta. En especial, la precisión requerida para el OCR y su tolerancia dependerá de los factores anteriormente mencionados, no obstante, si la naturaleza o importancia de la información consultada corresponde específicamente a datos críticos o legalmente vinculantes de carácter precontractual-contractual-postcontractual (valores, entidades, objetos, productos, fechas, alcance, áreas de estudio, entre otros), la precisión requerida será alta (95%-100%), esto con el fin de evitar resultados de búsquedas erróneos que

puedan tener consecuencias negativas para la Entidad. En el caso que la precisión del OCR sea media o baja, debido a los factores anteriormente mencionados, u otros, se sugiere que sea advertido el usuario de la herramienta o solución del nivel de precisión. La elección de las estrategias para mejorar la precisión del OCR recae en los interesados, quienes deben considerar la necesidad y los requerimientos técnicos específicos de la VT, las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado y los aspectos técnicos de los modelos.

Asimismo, los modelos del software/herramienta/aplicación deben ser capaces de procesar información y consultas que incluyan, como mínimo, palabras y expresiones en inglés. No obstante, se sugiere a los interesados reconocer y familiarizarse con las principales palabras o expresiones en otros idiomas que se utilizan habitualmente en el contexto técnico que rodea la información del proyecto.

7. La ANH está abierta a cualquier tipo de solución en la nube?

Respuesta: Se sugiere que la solución que se adquiera sea On-premise y Standalone, de tal forma que se pueda utilizar en otros procesos y no tenga una limitación por tiempo de servicio.

8. En cuanto a las capacitaciones mencionadas en el documento, se tiene la cantidad de personas, como sus perfiles a ser capacitados?

Respuesta: Se sugiere ver el apartado “PERSONAL MINIMO” del sondeo de mercado

9. La ANH plantea hacer una sesión técnica para aclarar los alcances?

Respuesta: En caso de realizarse el proyecto, los proponentes podrán presentar sus inquietudes y observaciones frente a los pliegos que hagan parte de los procesos de selección y contratación.

10. Solicitamos a la entidad ampliar la información con respecto a los aspectos funcionales de la solución para tener en cuenta en los tiempos de implementación y costos.

Respuesta: El propósito principal de la indexación de la información es la búsqueda, acceso y consulta eficiente de información, esto mediante una herramienta de indexación y búsqueda que integre nuevas tecnologías de la información y ciencia de la computación. Ahora bien, se sugiere revisar la necesidad y los requerimientos técnicos específicos de la VT, al igual que las actividades a ejecutar que se encuentran especificadas en el sondeo de mercado.