



DIRECCIÓN TÉCNICA DE HIDROCARBUROS

**Contrato 187 de 2025:
CRISTHIAN BOLIVAR RIASCOS RODRIGUEZ**

Informe actividades abril de 2025

tercer pago

Bogotá D. C., 29 de abril del 2025

Contenido

1.	Actividades objeto contractual	2
1.1.	Objeto contractual	2
1.2.	Obligaciones específicas	2
2.	Desarrollo de obligaciones contractuales (actividades)	2
2.1	Contribuir en la implementación de algoritmos de aprendizaje automático (machine learning) para optimizar los procesos de extracción y transformación a formatos numéricos de la información de interés para la DTH almacenada en el Banco de Información Petrolera.	2
2.2	Participar en el proceso de anotación o etiquetado de datos que se requieran para ayudar a los algoritmos de aprendizaje automático a comprender y clasificar información.	4
2.3	Contribuir en la implementación de algoritmos y aplicaciones utilizadas para analizar, clasificar e interpretar imágenes.	4
2.4	Contribuir en la elaboración de un informe de avance semestral, y dos informes finales técnicos y de gestión del proyecto	4
2.5	Las demás actividades que se requieran para el cumplimiento del objeto contractual y estén acorde con la naturaleza del contrato.	4
2.6	Participar en actividades de divulgación y socialización de actividades y productos técnicos del área, en eventos donde participe el SGC.	5

1. Actividades objeto contractual

A continuación, se describe el objeto contractual del contrato **187 de 2025** y las obligaciones específicas (actividades).

1.1. Objeto contractual

Prestar servicios profesionales para implementar mecanismos de validación de la calidad información procesada para proponer mejoras en el rendimiento y optimización de metodologías de Transformación Digital.

1.2. Obligaciones específicas

- ✓ Contribuir en la implementación de algoritmos de aprendizaje automático (machine learning) para optimizar los procesos de extracción y transformación a formatos numéricos de la información de interés para la DTH almacenada en el Banco de Información Petrolera.
- ✓ Participar en el proceso de anotación o etiquetado de datos que se requieran para ayudar a los algoritmos de aprendizaje automático a comprender y clasificar información.
- ✓ Contribuir en la implementación de algoritmos y aplicaciones utilizadas para analizar, clasificar e interpretar imágenes.
- ✓ Contribuir en la elaboración de un informe de avance semestral, y dos informes finales técnicos y de gestión del proyecto.
- ✓ Las demás actividades que se requieran para el cumplimiento del objeto contractual y estén acorde con la naturaleza del contrato.
- ✓ Participar en actividades de divulgación y socialización de actividades y productos técnicos del área, en eventos donde participe el SGC.

2. Desarrollo de obligaciones contractuales (actividades)

A continuación, se describen las actividades desarrolladas durante el mes de marzo.

2.1 Contribuir en la implementación de algoritmos de aprendizaje automático (machine learning) para optimizar los procesos de extracción y transformación a formatos numéricos de la información de interés para la DTH almacenada en el Banco de Información Petrolera.

- Se llevo a cabo una actualización a la metodología de extracción de información litológica de descripciones de muestras de zanja. Para la actualización se compilaron dentro de paquetes individuales tres de los subprocesos de la metodología: el proceso de validación de la información de documentos con descripciones de muestras de zanja, el proceso de extracción de información litológica y el proceso de carga de información a la base de datos de búsqueda de descripciones de muestras de zanja. A continuación, se detallan las actualizaciones hechas.

Validación de información de documentos con descripciones de muestras de zanja: para este proceso se compiló en un paquete llamado *metadata_val*, un script de Python que ejecuta una interfaz gráfica en donde se lleva a cabo la validación de la información obtenida a partir de la identificación automática de documentos con descripciones de muestras de zanja. El paquete se ejecuta desde la línea de comandos de Windows tomando como argumentos de entrada una lista de nombre de pozos cuyos documentos se quieren validar, la base de datos de búsqueda de muestras de zanja y la lista de pozos registrada en el Banco de Información Petrolera. La ejecución de este paquete despliega la interfaz y permite hacer actualizaciones a la base de datos con la información validada.

Extracción de información litológica de descripciones de muestras de zanja: para este proceso se compiló en un paquete llamado *extraction* un script que se encarga de ejecutar el proceso de extracción de información litológica mediante la implementación de un modelo de clasificación de textos y un modelo de reconocimiento de entidades. El paquete se ejecuta a través de la línea de comandos de Windows y toma como argumentos de entrada una lista de pozos sobre los que se ejecutará la extracción de información, la base de datos de búsqueda de documentos con descripciones de muestras de zanja desde donde se seleccionará el documento sobre el que se hará la extracción y por último una ruta para almacenar los resultados. El resultado de la extracción son dos archivos de Excel; el archivo *00_results.xlsx* donde se presenta un resumen de los resultados y *01_df_intervals.xlsx* en el que se encuentra la información litológica extraída. Este último archivo es el que será sujeto de validación y corrección en una etapa posterior.

Actualización de base de datos de búsqueda de descripciones de muestras de zanja: para este proceso se compiló un paquete llamado *load_info*, un script que de acuerdo a los documentos seleccionados para hacer la extracción de descripciones actualiza la base de datos para registrar de dónde proviene la extracción. Adicionalmente la rutina copia los documentos en una nueva ruta y los renombra siguiendo la estructura predefinida. Este paquete también se ejecuta desde la línea de comandos de Windows.

Los tres paquetes se alojaron como subcarpetas en la siguiente ruta:

\\srv-ar1\compartido_DTH\2023\510. GRUPO INVES. HIDRO. CS\5.
PROYECTOS\1001589_tdgi\repositorio_contratistas\cristhian_riascos\c_muestras_zanja\c_project

2.2 Participar en el proceso de anotación o etiquetado de datos que se requieran para ayudar a los algoritmos de aprendizaje automático a comprender y clasificar información.

- Durante el transcurso del mes se crearon dos conjuntos de datos de validación. El primer de ellos corresponde a un conjunto de datos cuyo propósito es validar la metodología que hace reconocimiento automático del nombre del pozo al que hace referencia un documento a partir de analizar el nombre del archivo. Para este caso se seleccionaron 800 documentos de interés que debían ser abiertos y leídos por una persona que luego debía consignar el nombre del pozo al que correspondía el documento. El segundo conjunto de datos tiene como propósito entrenar un modelo de aprendizaje automático capaz de reconocer el nombre del pozo de acuerdo con el contenido del documento, para este objetivo se recopilamos 100 documentos, en 50 de ellos se conoce el nombre del pozo puesto que se pudo determinar de analizar el nombre del archivo y en 50 de ellos no se conoce el nombre del pozo al que hace referencia el documento.

2.3 Contribuir en la implementación de algoritmos y aplicaciones utilizadas para analizar, clasificar e interpretar imágenes.

No se realizó esta actividad en el período.

2.4 Contribuir en la elaboración de un informe de avance semestral, y dos informes finales técnicos y de gestión del proyecto

- Elaboración del informe mensual de actividades que servirá de insumo para el informe semestral.

2.5 Las demás actividades que se requieran para el cumplimiento del objeto contractual y estén acorde con la naturaleza del contrato.

- Se desarrolló en Python una metodología que identifica documentos de interés según su contenido y determina el nombre de pozo al que hace referencia el documento. Para este objetivo se escribieron un conjunto de rutinas que analizan por n-gramas los nombres de los archivos. Este análisis consiste en segmentar el nombre de un archivo en tokens que posteriormente son reducidos mediante lematización a su forma base o lema. Luego se forman combinaciones de uno, dos o hasta tres lemas (n-gramas) y se hace un análisis de los n-gramas más comunes en los nombres de archivos que contienen información útil. Este análisis facilita el proceso de identificación de documentos de interés, puesto que se puede asociar el contenido de un documento con los n-gramas que contiene el nombre del archivo.

En un paso posterior mediante el estudio de los n-gramas contenidos en el nombre de un archivo se puede buscar mediante el uso de expresiones regulares, palabras que hagan referencia a nombres de pozo para así determinar el nombre del pozo al que hace referencia un documento.

Finalmente, se llevó a cabo una prueba de esta metodología para buscar documentos dentro del repositorio institucional que contuvieran información de salinidades de pozo. Una vez se identificaron los posibles documentos candidatos a contener este tipo de información se extrajo el nombre del pozo al que hacían referencia, con esto se logró construir posteriormente un mapa con la ubicación de los pozos para los que se encontró este tipo de información.

Ruta de ubicación del proyecto en Python:

```
\\srv-ar1\compartido_DTH\2025\510. GRUPO INVES. HIDRO. CS\510.52  
PROYECTOS\1001589_tdgi\repositorio_contratistas\cbrr\filename_analysis
```

- Se escribió un documento cuya función es servir como manual para guiar el proceso de validación y corrección de la información de descripciones de muestras de zanja. En este se detallan todos los pasos que se deben seguir para llevar a cabo dos procesos; verificar la información de los documentos que contienen descripciones de muestras de zanja (nombre de pozo, rango de páginas, idioma de las descripciones y formato) puesto que esta información es el insumo para llevar a cabo la extracción, y validar y corregir la información litológica extraída automáticamente de descripciones de muestras de zanja puesto que esta información debe cumplir con estándares geológicos de calidad.

Dentro del documento se detallan cada uno de los pasos a seguir para validar la respectiva información y se presentan ejemplos de las acciones a seguir en los diferentes casos que se pueden presentar.

Ruta de ubicación del documento:

\\srv-ar1\compartido_DTH\2023\510. GRUPO INVES. HIDRO. CS\5.
PROYECTOS\1001589_tdgi\repositorio_contratistas\cristhian_riascos\c_muestras_zanja\documentacion\manual_validacion_correccion_info_dmz.docx

- Participación en reuniones presenciales y virtuales con el grupo de trabajo de Transformación Digital con el objetivo de desarrollar las actividades propias del contrato (Anexo 1).

2.6 Participar en actividades de divulgación y socialización de actividades y productos técnicos del área, en eventos donde participe el SGC.

No se realizó esta actividad en el período.

Anexos

Anexo 1. Evidencia de las reuniones virtuales y presenciales asistidas durante el mes de abril.

Calendar

Hoy

< >

Abril de 2025

Mes

DOM	LUN	MAR	MIÉ	JUE	VIE	SAB
30	31 4:30pm Identificación	1 de abr 3pm Extracción DMZ	2	3 11am (virtual) Manejo I 12:15pm Revisión extra	4 10am Revisión Extraacc	5
6	7 9am CO2 - lista pozos 10:30am Revisión proc 11:30am Conjunto de d 4pm Construcción cor	8	9	10 8am Revisión seguiri 11am Revisión protoco 2pm protocolo extract 3pm Revisión procedir	11 1pm Revisión protocol 4:15pm protocolo DMZ	12
13	14 2pm Revisión protocol	15 11:15am Revisión proto 2:30pm Continuación I	16 8am Actualización inte 11am Proceso de extra	17	18	19
20	21 9am Tareas ejecución I 3pm CO2 - classify do	22	23	24 9:30am Reunión DMZ I 10:30am culminación I 1:30pm Revisión protot 4:30pm revision protot	25 11:30am prtocolo dmz 4:30pm Cierre de prot	26
27	28 10:30am Revisión carg	29 11am Cierre protocolo 3pm Capacitación "co	30	1 de may	2	3

+ Crear

Abril de 2025

< >

D L M X J V S

30 31 1 2 3 4 5

6 7 8 9 10 11 12

13 14 15 16 17 18 19

20 21 22 23 24 25 26

27 28 29 30 1 2 3

4 5 6 7 8 9 10

Reunirse con...

Buscar a gente

Estadísticas de tiempo

1 - 30 DE ABR DE 2025

Tiempo en reuniones: 8 h

Más información

Mis calendarios

☒ Cristhian Bolívar Riascos ...
 ☐ Cumpleaños
 ☐ Tasks

Otros calendarios +

☐ Festivos en Colombia