**INFORME DE PROPUESTAS PARA LOS TALLERES DE WEB SCRAPING Y RECUPERACIÓN DE INFORMACIÓN DE TEXTO**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Practicante** | **Web Scraping** | | **Recuperación de información de texto** | | **Comentario** |
| **Viabilidad** | **Justificación** | **Viabilidad** | **Justificación** |
| Luis Egúsquiza | Sí es viable | Sí justifica | Sí es viable | Sí justifica | Las dos propuestas son viables. Es importante definir el método de extracción de documentos de tesis para la aplicación de recuperación de información de texto. Una posible alternativa es implementar una solución de Web Scraping enfocada en recolectar datos directamente del repositorio de tesis, en lugar de centrarse en información procedente del MTC. |
| Diego Gonzáles | Sí es viable | Sí justifica | Sí es viable | Sí justifica | Las dos propuestas son viables. En el caso de la aplicación de Recuperación de Información de Texto, resulta crucial determinar con precisión la información específica a extraer e identificar los patrones en los que esta se organiza. |
| Jherson Cueva | Sí es viable | Sí justifica | Sí es viable | Sí justifica | Las dos propuestas son viables. Para la aplicación de Recuperación de Información de Texto, es importante considerar qué información se desea extraer exactamente, e ir identificando patrones en los que dicha información se organiza. |
| Pamela Salazar | Sí es viable | Sí justifica | Viable, pero difícil | Sí justifica | La aplicación de Web Scraping es viable. Sin embargo, la viabilidad de la aplicación para Recuperación de Información de Texto podría verse limitada debido a que las tablas en los informes del INEI carecen de bordes y la información parece organizarse de manera no estructurada. Aunque es factible extraer esta información, el proceso podría requerir una inversión significativa de tiempo y esfuerzo. |
| Michael Encalada | Sí es viable | Sí justifica | Sí es viable | Sí justifica | Las dos propuestas son viables. Sin embargo, con respecto a la propuesta de Web Scraping, es importante considerar que existen maneras más sencillas de acceder a dicha información. Sin embargo, podría ser un bien ejercicio realizar esta aplicación. |
| Irinia Vargas | Sí es viable | Sí justifica | Sí son viables | Sí justifica | Las dos propuestas son viables. En el caso de la propuesta de Web Scraping la aplicación resulta interesante, porque se puede generar un proceso automatizado de consulta para extraer tablas dados los parámetros correspondientes “Región”, “Provincia”, etc. Respecto a las propuestas de Recuperación de Información de Texto, la aplicación para Evaluación Docente muestra una mayor viabilidad y una optimización más efectiva del costo-beneficio. |
| Paola Aranda | Sí es viable | Sí justifica | Sí es viable | Sí justifica | Las dos propuestas son viables. Respecto a la aplicación de Web Scraping, considerando que el portal web de Transparencia incluye datos de diversas entidades estatales, existe el potencial de desarrollar un scraper más generalizado. Este no se limitaría solo a la información de QaliWarma, sino que también podría acceder a los datos de otras entidades presentes en el portal. |
| Esteban Cabrera | Sí es viable | Sí justifica | Sí es viable | Sí justifica | Las dos propuestas son viables. Con respecto a la aplicación de Recuperación de información de texto, es importante considerar que los documentos están escaneados, y que en este caso sí es imprescindible aplicar una librería de interpretación de imagen como Pytesseract. |