Submission code: 9ZWAN0XPZN

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

WE-ME: An ETL Tool for ANP Well Data to Convert Messy Files into Structured Information

Maria Eduarda Candida Francklim Francisco (Federal University of Rio de Janeiro), Francisco Manoel Tognoli (Federal University of Rio de Janeiro)

WE-ME: An ETL Tool for ANP Well Data to Convert Messy Files into Structured Information

Please, do not insert author names in your submission PDF file

Introduction

Brazil hosts one of the largest well data repositories in the world, managed by the National Agency of Petroleum, Natural Gas, and Biofuels (ANP). The revitalization programs for exploration and production in onshore (REATE) and offshore areas (PROMAR) have made these data more accessible, opening new possibilities for exploration and research. However, the vast volume and diverse file formats present significant challenges for data preparation, especially if linked with advanced analyses in seismic interpretation software. The absence of standardization within historical records and between them and contemporary ones makes manual processing difficult, making it slow and prone to errors.

Method and/or Theory

To address these issues, a system named WE-ME (Well data Extraction - Made Easy) was developed to automate the organization and processing of ANP data. It filters and structures essential information, including UTM coordinates, datum, directional data (measured depth, true vertical depth, inclination, azimuth), lithological information, lithostratigraphy, chronostratigraphy, checkshots, evidence of hydrocarbons and the year the wells were drilled.

The system was implemented in Python, using the os, unicodedata, re, pandas, and XlsxWriter modules. The os module provided a platform-independent interface to interact with the file system, enabling navigation, access, and creation of directories and files. The unicodedata module was used to normalize and clean strings by decomposing accented characters and removing non-ASCII symbols such as "ç". The re module enabled the use of regular expressions to search, extract, and manipulate specific patterns in the unstructured .txt files. The pandas library offered powerful data structures (DataFrames) that allowed for efficient tabular processing, merging, and grouping of information from multiple sources. Finally, XlsxWriter was used to generate the final Excel spreadsheets, with customized formatting for columns and multiple pages.

The validation step used data from 343 wells from the Santos and Campos Basins (offshore), and Potiguar and Acre Basins (onshore), demonstrating successful integration with a seismic interpretation software.

Results and Conclusions

The outputs are summarized spreadsheets and files ready for integration with seismic interpretation software. By leveraging Python's flexibility, WE-ME enables the creation of customizable scripts that streamline repetitive tasks, improve data accessibility and facilitate reproducibility. Processes to extract, transform and load (ETL) data, which used to take weeks, can now be performed in minutes. This work highlights how automation can simplify the handling of big data, stimulating the use of the large Brazilian well database in exploratory studies and contributing to the optimization in geosciences and the energy industry. Furthermore, the success of this approach in the Brazilian context highlights its potential applicability to other regions rich in resources - both natural and data-related - that face similar challenges.