

Multi-source Question Answering Framework Using Transformers for Attribute Extraction

Authors : Prashanth Pillai, Purnaprajna Mangsuli

Abstract : Oil exploration and production companies invest considerable time and efforts to extract essential well attributes (like well status, surface, and target coordinates, wellbore depths, event timelines, etc.) from unstructured data sources like technical reports, which are often non-standardized, multimodal, and highly domain-specific by nature. It is also important to consider the context when extracting attribute values from reports that contain information on multiple wells/wellbores. Moreover, semantically similar information may often be depicted in different data syntax representations across multiple pages and document sources. We propose a hierarchical multi-source fact extraction workflow based on a deep learning framework to extract essential well attributes at scale. An information retrieval module based on the transformer architecture was used to rank relevant pages in a document source utilizing the page image embeddings and semantic text embeddings. A question answering framework utilizing LayoutLM transformer was used to extract attribute-value pairs incorporating the text semantics and layout information from top relevant pages in a document. To better handle context while dealing with multi-well reports, we incorporate a dynamic query generation module to resolve ambiguities. The extracted attribute information from various pages and documents are standardized to a common representation using a parser module to facilitate information comparison and aggregation. Finally, we use a probabilistic approach to fuse information extracted from multiple sources into a coherent well record. The applicability of the proposed approach and related performance was studied on several real-life well technical reports.

Keywords : natural language processing, deep learning, transformers, information retrieval

Conference Title : ICDAR 2022 : International Conference on Document Analysis and Recognition

Conference Location : Jerusalem, Israel

Conference Dates : November 29-30, 2022