

The CDA Challenge: Unstructured Data Mining

Many organisations struggle with extracting the information stored within unstructured datasets like Word, PDF, Log files and images. IDS took up the challenge set out by CDA to find value in unstructured data collected from 11,000 wells over the last 50 years. This data consisted of documents in a multitude of formats; from scanned typewritten documents, TIFF images and non-digitised logs.

From the outset, the IDS project team agreed that all of these document types should be searchable and utilise a web-based interface. The solution was built on top of the incredibly powerful Elastic Stack, meaning the data was searchable once ingested into the search engine, Elasticsearch.

IDS presented the outcome of the project at a conference hosted by CDA. The end product is delivered in a portal whereby a user can enter an industry known term with the results being returned in a graphical interface. For example – a search for “stuck pipe” would search all documents for this keyword then display the related wells via a split screen with a map beside the related well header data.

Due to the success of the CDA challenge, IDS have incorporated the technologies and learnings in the improved Lessons Learned solution which now allows all lessons entered at any regional location to be searchable at a global level.

