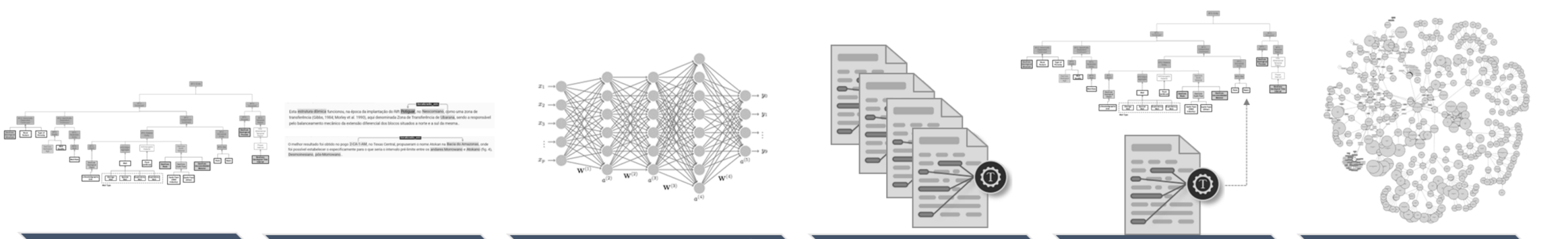
PETRO KGRAPH: A METHODOLOGY FOR EXTRACTING KNOWLEDGE GRAPH FROM TECHNICAL DOCUMENTS

AN APPLICATION IN A CORPUS FROM THE OIL AND GAS INDUSTRY.

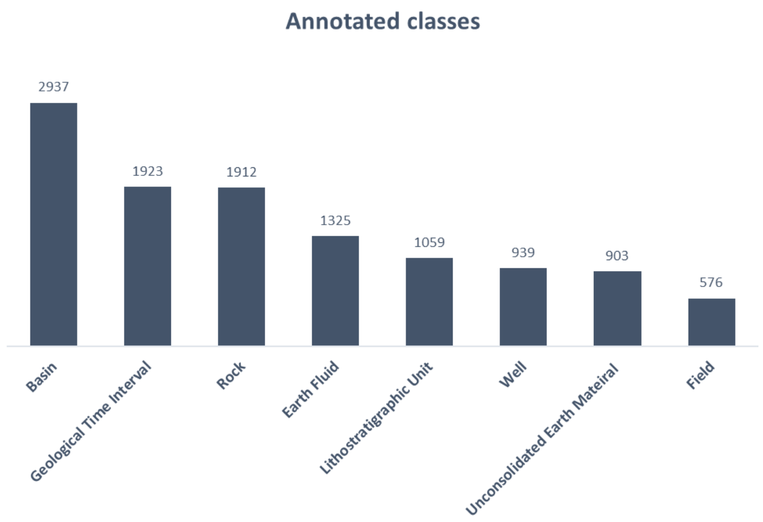
FÁBIO CORRÊA CORDEIRO

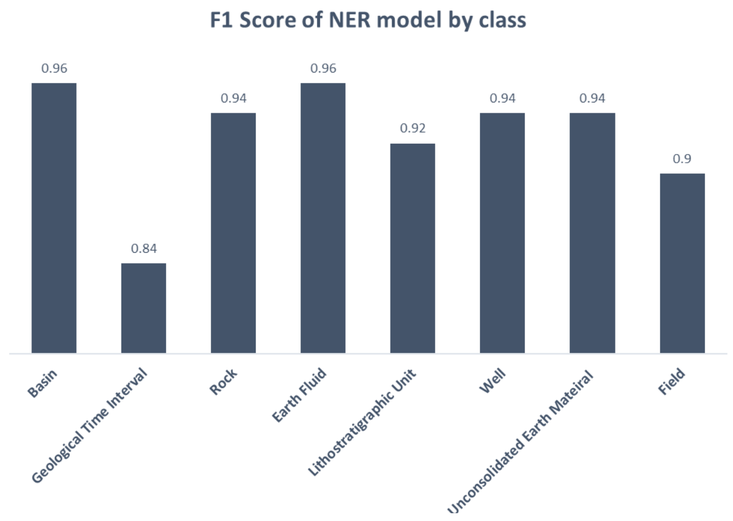
INTRODUCTION

Extracting information from huge datasets is the basis of many recent technologic developments and has become an important research topic. Many companies have shown interest in gathering strategic information from their repositories of documents. It is particularly relevant for the oil and gas industry because they have large repositories with geoscientific reports from several decades of production. We proposed a methodology for extracting entities and relations from domain-specific technical documents to populate a knowledge graph. We believe that a specialized knowledge graph can improve the results of domain-specific search engines.



ANNOTATION AND MODEL





PETROLÊS

Petrolês is a public repository of artifacts for Natural Language Processing applications in the petroleum domain in Portuguese.

This repository aims to serve as a reference for artificial intelligence research groups and companies related to the oil and gas sector.

Access using this QR Code.

PARTIAL RESULTS

We had achieved some preliminary results and materialized them in the following paper and patent.

Paper or patent	Journal, conference or institution	Link
Petrolês - How to Build a Specialized Oil and Gas Corpus in Portuguese	Rio Oil and Gas Expo and Conference 2020	
Portuguese word embeddings for the oil and gas industry: Development and evaluation	Computer and Industry journal	
Embeddings for Named Entity Recognition in Geoscience Portuguese Literature	Language Resources and Evaluation Conference (LREC 2020)	
Method for extracting and structuring information	Brazilian National Institute of Industrial Property (INPI)	Confidentiality period
Improvement Optical Character Recognition for Structured Documents using Generative Adversarial Networks	21st International Conference on Computational Science and Its Applications (ICCSA 2021)	

PETROBRAS
FGV EMAp

Advisor: Flávio Codeço Coelho
Coadvisor: Renato Rocha Souza

fabio.cordeiro@petrobras.com.br
linkedin.com/in/fabio-correa-cordeiro
github.com/fabiocorreacordeiro
github.com/Petroles