

Optimizing Application Landscapes: A Sentence Similarity Approach to Rationalizing Subsurface Software

Abdullahi Yusuf¹, Dhiyab Al Abri¹, Naima Abri¹

¹Petroleum Development Oman

Abstract

This paper presents an innovative approach for optimizing application landscapes in the oil and gas industry, employing fine-tuned sentence similarity algorithms to rationalize subsurface software portfolios. We developed a unique framework that integrates linguistic analysis with industry-specific nuances, addressing gaps identified in current methodologies. Through a meticulous process of data collection, preprocessing, and algorithm selection, we tailored our approach to effectively process technical language specific to the oil and gas sector.

The results from our experiments, which involved rigorous testing and training phases, demonstrated a high degree of accuracy in identifying redundant and overlapping applications, with a ~7% increase in Spearman Correlation post fine-tuning. A case study application in a real-world scenario illustrated the practical benefits of our method, showing a potential of up to ~25% reduction in software redundancies and an improvement in IT resource allocation a multimillion-dollar savings for mature enterprise. Our approach not only enhances the efficiency of software utilization but also sets a new precedent in applying natural language processing techniques for IT portfolio management in the oil and gas industry.

Innovation

At the heart of our research is the innovative application of natural language processing (NLP) techniques to a domain traditionally dominated by manual and heuristic-based approaches. Our framework represents a pioneering effort to harness the power of linguistic analysis and machine learning for the optimization of IT portfolios in the oil and gas industry.

The novelty of our approach lies in its fine-tuned sentence similarity algorithms, which have been meticulously adapted to grasp the complex technical language inherent to the oil and gas sector using a distilled Transformer model fine-tuned on GPT4 output. This level of customization is unprecedented and addresses a critical gap in existing methodologies, which often overlook the unique linguistic challenges presented by industry-specific terminologies.

Our experiments showcase the effectiveness of our approach, with a ~7% increase in Spearman Correlation post fine-tuning, indicating a high degree of accuracy in identifying software redundancies. The broader implications of our research extend beyond the oil and gas industry. By demonstrating the viability of applying advanced NLP techniques to IT portfolio management, our work opens new avenues for research and practice. It sets a precedent for how linguistic and machine learning algorithms can be leveraged to solve complex business challenges, paving the way for further innovations in this field.