T5-deep learning techniques have significantly impacted the field of semantic text matching . existing state-of-the-art models are primarily tailored for short documents like tweets . this article introduces CoLDE, a Contrastive Long Document Encoder framework based on transformers . the framework enables interpretable comparisons by providing fine-grained similarity scores .

Bart-

Supervised Contrastive Learning for Interpretable Long-Form Document Matching. CoLDE utilizes unique positional embeddings and a multi-headed chunkwise attention layer within a supervised contrastive learning framework to capture similarity at various levels within and across documents. The framework enables interpretable comparisons by providing fine-grained similarity scores. Evaluation on three long document datasets demonstrates CoL DE's superior performance in document matching, robustness to document length variations and text perturbations, and provision of interpretable results. The proposed model's code is publicly available. It is designed to address the challenges associated with handling long documents such as scientific papers, legal documents, and patents. It uses a Contrastive Long Document Encoder framework based on transformers, designed to solve the problem of long-form document matching.

Gemini- CoLDE (Contrastive Long Document Encoder) is a novel framework designed for semantic matching of long-form documents. It leverages supervised contrastive learning and incorporates specialized positional embeddings and a multi-headed chunkwise attention layer. Unlike existing models tailored for short documents, CoLDE effectively captures similarity at various levels within and across long documents. Evaluations on benchmark datasets demonstrate its superior performance, robustness to document length variations and text perturbations, and ability to provide interpretable similarity scores, making it a valuable tool for long-form document matching tasks.

Gpt-4 The article discusses the development of CoLDE (Contrastive Long Document Encoder), a novel framework designed to enhance the performance of semantic text matching (STM) for long-form documents such as scientific papers, legal documents, and patents. Traditional STM models, which excel with shorter texts like tweets and user reviews, struggle with the complexity and length of longer documents. CoLDE addresses this gap by incorporating advanced transformer-based techniques specifically adapted for longer texts.

CoLDE integrates unique positional embeddings and a multi-headed chunkwise attention layer within a supervised contrastive learning framework. This design choice is pivotal as it allows the model to effectively manage and interpret the extensive information contained in long documents. By focusing on both intra-document and inter-document similarities, CoLDE can analyze and compare extensive texts at various levels, providing a more nuanced understanding of document content.

The framework's use of supervised contrastive learning is particularly noteworthy. This approach not only enhances the model's ability to discern relevant features in the text but also facilitates the generation of interpretable outputs. Users can obtain fine-grained similarity scores that offer clear insights into the basis of the model's comparisons, making CoLDE a valuable tool for applications requiring detailed document analysis and understanding.

Evaluations of CoLDE on three different long document datasets have shown its effectiveness. The results highlight the model's superior performance in matching documents accurately, its robustness against variations in document length and textual perturbations, and its ability to provide interpretable insights into the similarities and differences between documents. The availability of CoLDE's code to the public further underscores its potential utility and impact in the field of document analysis and semantic matching, promising significant advancements in how complex documents are processed and understood in various professional and academic contexts.

Gp-3 The field of semantic text matching (STM) has seen significant progress in deep learning techniques, but existing models are limited to short documents, hindering their application to long-form texts like scientific papers and legal documents. To address this gap, the article introduces CoLDE, a Contrastive Long Document Encoder framework based on transformers. CoLDE incorporates unique positional embeddings and a multi-headed chunk wise attention layer in a supervised contrastive learning setup to capture similarities within and across documents at different levels. This approach allows for interpretable comparisons through fine-grained similarity scores. Evaluation on three long document datasets showcases CoLDE's superior performance in document matching, resilience to document length variations and text alterations, and provision of interpretable outcomes. Notably, the model's code is openly accessible, promoting transparency and reproducibility in research. CoLDE's innovative design offers a promising solution for handling long-form documents, expanding the applicability of deep learning techniques in text matching tasks.