**Text Generation from Knowledge Graphs with Graph Transformers**

1. **Kiruthika. B - Register No: 2034025**
2. **Rithanya. R.B - Register No: 2034040**
3. **Sasruthisri. K - Register No: 2034042**



November 2023

**DEPARTMENT OF COMPUTING (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

**COIMBATORE INSTITUTE OF TECHNOLOGY**

**(Autonomous Institution affiliated to Anna University)**

**COIMBATORE – 641014**

**Graph Analytics Lab**

**ABSTRACT:**

The proliferation of data in knowledge graphs presents a unique challenge in natural language processing: generating coherent texts that span multiple sentences and convey complex ideas. Traditional text generation techniques struggle with the non-hierarchical nature and structural variety of knowledge graphs.

Recent research has made strides in addressing these issues, yet the generation of multi-sentence texts from graphical data remains a significant hurdle. Current methods often fail to capture the rich relational structure and variety inherent in knowledge graphs, leading to less informative and structurally weak texts.

This project introduces a novel graph-transforming encoder that respects the relational structure of knowledge graphs without the need for linearization or hierarchical organization. By incorporating this encoder into an end-to-end trainable encoder-decoder setup, we can generate texts from knowledge graphs with improved informativeness and document structure.

Our proposed methodology leverages the power of graph transformers to address the challenges posed by knowledge graph-based text generation. The results from automatic and human evaluations indicate that our approach outperforms existing methods, producing texts that are not only more informative but also exhibit better document structure.

**Keywords:**Knowledge Graphs, Text Generation, Graph Transformers, Natural Language Processing, Encoder-Decoder Models.

**PROBLEM STATEMENT:**

Despite the vast amount of information encapsulated within knowledge graphs, the challenge remains to effectively generate coherent and structured multi-sentence text. Current methods often fail to fully utilize the relational structure of knowledge graphs, resulting in texts that lack informativeness and structural integrity. This project aims to address these limitations by introducing a novel graph-transforming encoder capable of maintaining the relational data integrity of knowledge graphs while generating informative and well-structured text.

**INTRODUCTION:**The project at hand introduces GraphWriter, a groundbreaking graph-transforming encoder designed to harness the relational structure of knowledge graphs without the constraints of linearization or hierarchical organization. This innovation allows for a more structured representation of content, setting it apart from traditional text generation methods.

GraphWriter’s new attention model for graph encoding is a testament to its advanced capabilities in handling complex data structures. Additionally, the project contributes the AGENDA dataset to the text generation community, serving as a robust resource for model development and evaluation.

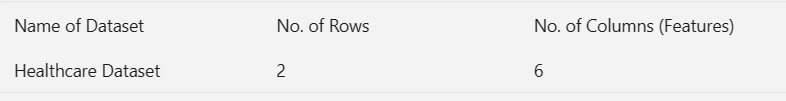
Through both human and automatic evaluations, GraphWriter has demonstrated its superiority over strong baselines, effectively generating coherent multi-sentence texts. Despite facing challenges such as repetition and entity coverage, GraphWriter’s success underscores the critical role of integrating knowledge graphs into text generation systems, promising enhanced performance across various domains.

The project’s key contributions are noteworthy: the proposal of a graph transformer encoder, the innovative use of information extraction output as a graph for attention-based encoders, and the provision of a dataset pairing knowledge graphs with scientific texts. These advancements pave the way for future research and applications in the field of natural language processing.

**APPROACH:**

1. [**Graph Transformation**: GraphWriter utilizes a novel graph-transforming encoder that can leverage the relational structure of knowledge graphs without imposing linearization or hierarchical constraints1](https://arxiv.org/pdf/1904.02342.pdf).
2. [**Attention Model**: It incorporates a new attention model for graph encoding, allowing for a more structured representation of content compared to traditional methods1](https://arxiv.org/pdf/1904.02342.pdf).
3. [**Dataset Introduction**: The AGENDA dataset is introduced, providing a valuable resource for the text generation community to further develop and evaluate their models1](https://arxiv.org/pdf/1904.02342.pdf).
4. [**Evaluations**: Human and automatic evaluations demonstrate the utility of GraphWriter compared to strong baselines, highlighting its effectiveness in generating coherent multi-sentence texts1](https://arxiv.org/pdf/1904.02342.pdf).
5. **Challenges and Contributions**: While GraphWriter shows promising results, it also acknowledges challenges such as repetition and entity coverage that need to be addressed. [The key contributions include proposing a graph transformer encoder, demonstrating the use of information extraction output as a graph for attention-based encoders, and providing a dataset of knowledge graphs paired with scientific texts](https://arxiv.org/pdf/1904.02342.pdf)

**DATASET**



**Features and Descriptions:**

1. **Title**: The title of the dataset entry, which gives a brief idea of the content. For example, “Improving Patient Care Through Electronic Health Records” and “Advancements in Telemedicine Technology”.
2. **Entities**: Key terms or concepts within the dataset entry, such as “electronic health records”, “patient care”, “healthcare providers”, “medical history”, “treatment plans”, and “diagnostic tests”.
3. **Types**: The category of each entity, represented in a sequence corresponding to the entities. For instance, “<material> <task> <material> <material> <material> <material>”.
4. **Relations**: The relationships between entities, showing how one entity affects or is related to another. Examples include “electronic health records – IMPROVE – patient care” and “healthcare providers – ACCESS – electronic health records”.
5. **Abstract**: A summary of the study or dataset entry that provides an overview of the topic and its significance.
6. **Abstract\_og**: The original abstract text, which may be used for comparison or reference.