Content Reconstruction: The Evolution of Texts through Semantic Networks and LLMs

Dmytro Lande, Oleh Humeniuk

Igor Sikorsky Kyiv Polytechnic Institute

Abstract

This paper presents a novel methodology for generating texts with new meanings through the modification of semantic networks constructed using large language models (LLMs). The approach involves building a semantic network from an initial text, modifying this network by adding and removing nodes and connections, and then generating new texts based on the modified network. The paper outlines the mathematical model, application scenarios, prompts for implementation, and methods for evaluating the generated texts.

Keywords: Large language models; Text generation; AI content creation; Semantic networking; Text reconstruction

Introduction

With the advancement of artificial intelligence (AI), particularly large language models (LLMs) like GPT-4, new approaches to automated text generation have emerged. These technologies find applications in various domains, from content creation and marketing to scientific research and data analysis. However, their potential is far from fully realized. Specifically, generating texts with new meanings through targeted modification of semantic networks presents a promising direction, offering the ability to create high-quality content tailored to specific tasks and audiences.

The proposed method not only generates texts based on existing data but also significantly alters their content, making it a unique tool in the AI context.

Content modification, for instance, for adaptation to different user categories, can be applied in advertising, product promotion, education, and other fields.

The proposed method can be likened to genetic modification and executable program modification, where alterations lead to significant changes in outcomes. In genetics, DNA modification involves altering genetic material to achieve desired traits in an organism or its offspring. Similarly, modifying semantic networks can lead to substantial changes in the generated texts.

In the context of executable programs, disassembly, code modification, and reassembly are analogous to modifying semantic networks. Disassembly involves breaking down a program into its component parts, modifying the code, and then reassembling it to achieve new functionalities or improve performance.

Analogies and Background

The proposed method is analogous to genetic modification and executable program modification:

Genetic Modification: Genetic modification involves altering DNA to produce desired traits in organisms or their offspring [1]. Similarly, modifying semantic networks involves changing nodes and connections to generate texts with new meanings. This process can lead to significant transformations in the output, akin to how genetic modifications can affect an organism’s characteristics.

Executable Program Modification: In software engineering, disassembly, code modification, and reassembly are used to alter the functionality of programs [2]. This process parallels modifying semantic networks—disassembling the network, making adjustments, and reassembling it to produce new results.

Classic Text Generation Models

Early approaches to text generation included:

Markov Models: Used for generating text based on probabilistic chains, predicting the next word or phrase based on previous ones [3].

Recurrent Neural Networks (RNNs): Considered previous context and word sequences for generating new texts, improving the coherence and quality of generated content [4].

Large Language Models

Large language models like GPT-4 are powerful tools capable of analyzing and generating text based on vast amounts of input data. These models can understand complex contexts and dependencies, making them invaluable for creating coherent and contextually relevant content [5].

Transformers, such as GPT models, revolutionized text generation by using attention mechanisms to analyze and understand the context of entire sentences rather than just individual parts.

LLMs excel at generating texts that maintain the stylistic and substantive characteristics of the original material. This capability opens up numerous possibilities for applications in marketing, content development, analytical research, and beyond.

With the advent of large language models, a new era in text generation began. Models like GPT-4 demonstrate the ability to analyze large volumes of text data and generate content that retains the stylistic and substantive integrity of the original material.

Additionally, the work by D. Lande and L. Strashnoy [6] has made significant contributions to the development of text analysis methods and semantic networks, reflecting the current advancements in semantic networking approaches.

Methodology

Constructing the Semantic Network

The initial step involves constructing a semantic network based on the input text. This is achieved using the following prompt:

Prompt 1: "Highlight pairs of the most related concepts from the given text."

This prompt is applied multiple times to the same text to gather various concepts and connections. The frequency of these concepts and connections determines their weight in the network.

Mathematical Formalism

Node Weight:

$$w_i = \alpha \cdot Frequenci(c_i) + ExpertEvalution$$

Connection Weight:

$$w_{ij} = \beta \cdot Frequenci(c_i, c_j) + ExpertEvalution$$

where w_i is the weight of node c_i, w_{ij} is the weight of the connection between nodes c_i and c_j, and α, β is the normalization factor.

Swarm of Virtual Experts

The concept of a "swarm of virtual experts" refers to using multiple queries to gather insights on concepts and connections, with each query representing the work of a virtual expert. The aggregate responses from numerous queries, each with designated roles, simulate the work of a large number of experts. This approach aligns with the principles outlined by D. Lande and L. Strashnoy in their work on semantic networking [6].

Network Modification

Modification of the network involves: Adding Nodes and Connections, Removing Nodes and Connections & Weight Adjustments.

Adding Nodes and Connections: New nodes and connections can be added to the semantic network, expanding its structure and enhancing its ability to generate new meanings.

Adding Nodes:

$$N_{new} = N_{original} \cup \{n_{new}\}.$$

Adding Connections:

$$E_{new} = E_{original} \cup \{n_i, n_j\}.$$

Removing Nodes and Connections: Nodes and connections can be removed either automatically using LLMs or manually by experts. This optimization process can be likened to genetic modification in biology, where removal or alteration of DNA segments can lead to significant changes in the organism or its offspring.

Removing Nodes:

$$N_{updated} = N_{original} \setminus \{n_{remove}\}.$$

Removing Connections:

$$E_{updated} = E_{original} \setminus \{n_i, n_j\}.$$

Weight Adjustments: The weight of each node and connection is determined based on its frequency of occurrence and expert evaluations.

Text Generation

Once the semantic network is modified, it is used by LLMs to generate new texts. Example prompts for this stage include:

Prompt 2: "Create a new text based on the provided semantic network, preserving the logical connections and stylistic features of the original text."

Prompt 3: "Generate a text based on this network, but adapt it for a specific audience (e.g., business, youth)."

Methodology Variations

Possible variations of the method include:

Reconstructing the Semantic Network from Generated Texts: This approach allows for comparison between the original and generated networks to assess the extent of changes, analogous to modifying executable programs through disassembly, code modification, and reassembly.

Measuring Network Similarity: Comparing the original and modified networks in terms of node and connection similarity.

Node Similarity:

$$Node\ Simlarity = \frac{|V_{original}|}{|V_{original} \cup V_{generated}|}100\%.$$

Connection Similarity:

$$Connection\ Simlarity = \frac{|E_{original}|}{|E_{original} \cup E_{generated}|}100\%.$$

For network comparison, use the following prompt:

Prompt 4: "Compare two semantic networks and determine the percentage similarity of nodes and connections between them."

Evaluation of Methodology

As previously noted, evaluating the quality of generated texts involves comparing their semantic networks with the original ones. This quantitative assessment helps to understand how effectively the network modification method worked.

Example

The original message is being reviewed:

If you had your way with Android 14 and you're ready to move on to the next one, Google has your back and Android 15 is coming down the way. For users already enrolled in the QPR beta program on a compatible Pixel device, the first Android 15 QPR beta update has likely been offered. If opting out of the beta program without wiping the device is preferred, it's advised to avoid installing the beta. Instead, users should wait for the official release version update to become available on their Pixel device. After applying the stable release update, they can opt out without a data wipe, provided they do not install the subsequent beta update.

As a result of querying the LLM with prompt "**Provide 25 pairs of related concepts from the text in a numbered list format as "concept1;concept2". For example, "Android;Google"**", we obtain pairs of concepts corresponding to the semantic network (Fig. 1):

Android;Google Android 14;Android 15 Android 15;QPR beta update QPR beta update;Pixel device Pixel device;Google Android 14;beta program

Android 15;official release
official release;Pixel device
QPR beta program;stable release
stable release;beta update
Pixel device;beta update
QPR beta program;Google
Android 14;Pixel device
Android 15;stable release
QPR beta update;Android 14
beta program;stable release
beta program;Google
Google;stable release
beta program;Android 15
QPR beta update;official release
Pixel device;beta program
Android 15;Google
QPR beta program;Pixel device
official release;beta program
stable release;Pixel device

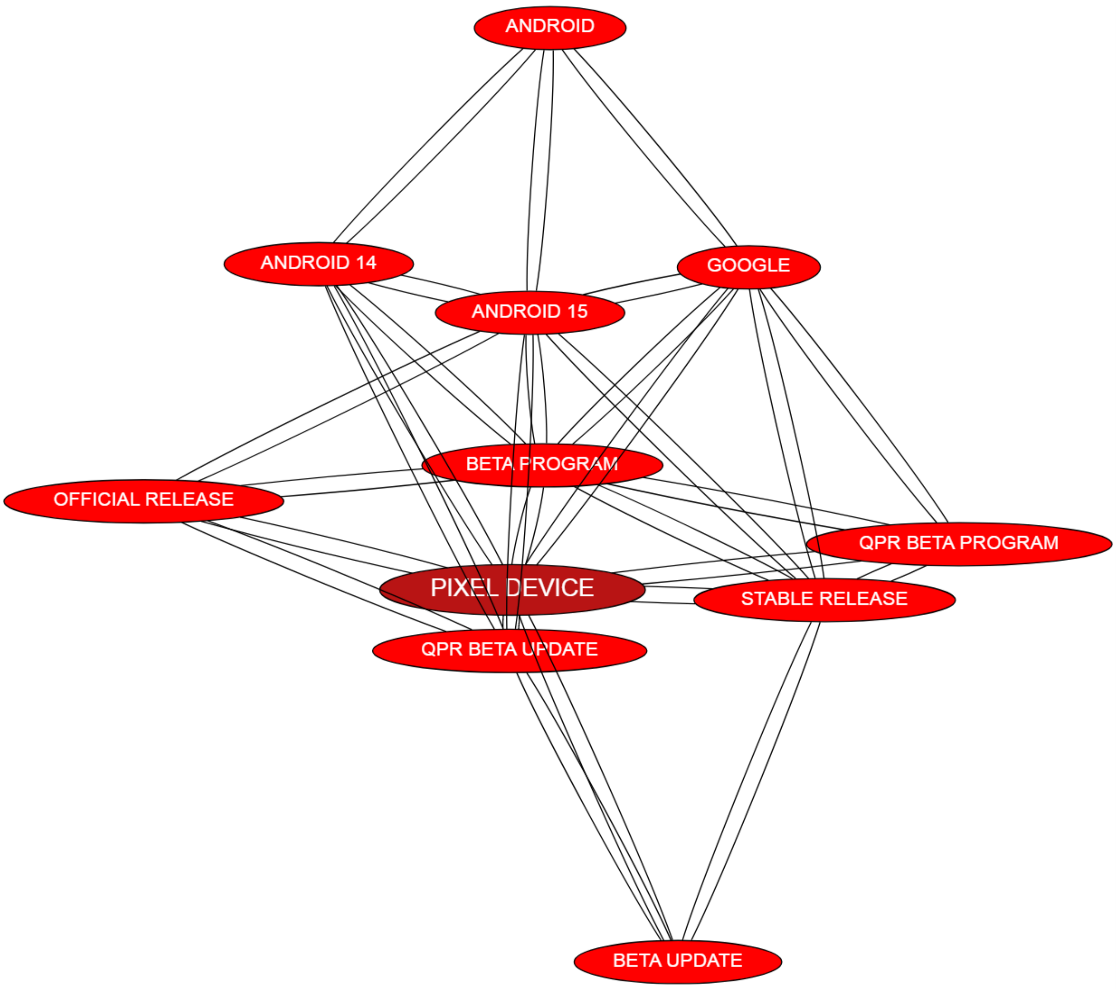

Figure 1. Core of the Semantic Network of the Original Text

Adding new nodes and connections related to the medical field to the original network (Fig. 2):

Medicine; Android 15 Medicine; treatment of people Treatment of people; Google Android 15; Convenient for doctors Treatment of people; stable release

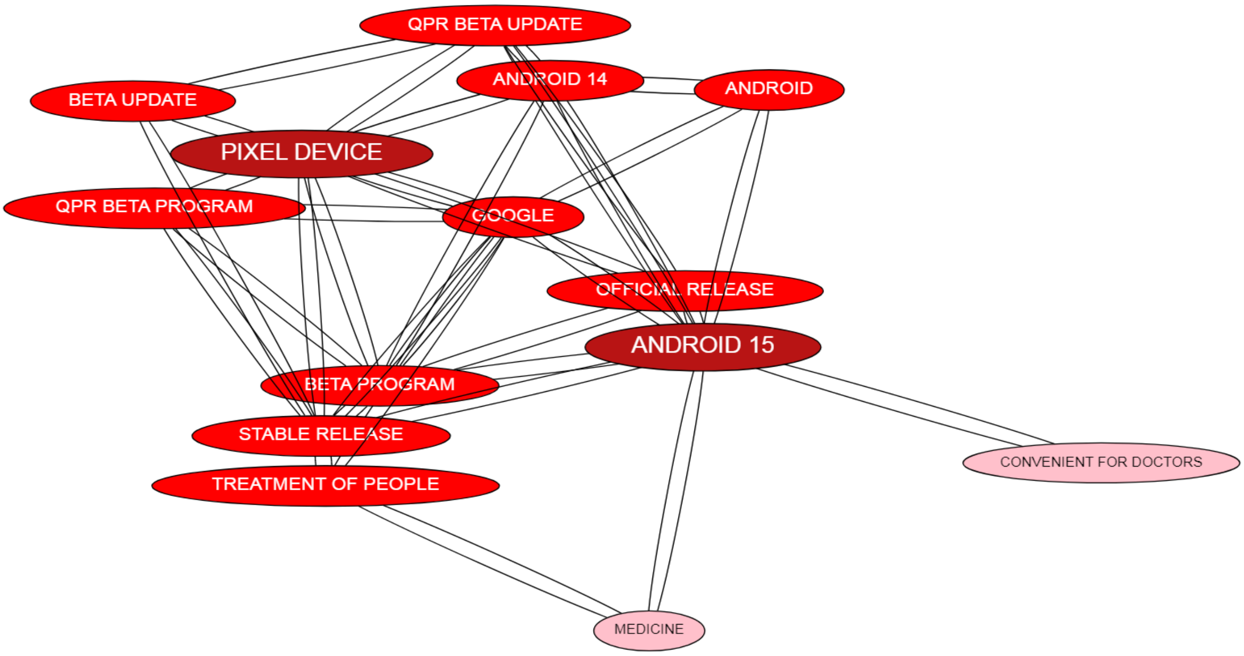

Figure 2. Core of the Modified Semantic Network

We're getting a new message:

Google's ongoing development in the Android ecosystem is paving the way for Android 15, which will succeed Android 14. Users who are part of the QPR beta program on compatible Pixel devices may have already received the Android 15 QPR beta update. For those wishing to avoid data loss, it's advisable to skip the beta update and wait for the official release. Once Android 15's stable release is available, users can exit the beta program without data wipe if they do not install any additional beta updates. In a different context, Android 15 is also finding applications in the medical field. Its features are designed to be convenient for doctors and assist in the treatment of patients. The stable release of Android 15 might offer new tools and enhancements that could be beneficial in medical settings, highlighting Google's commitment to integrating technology across various domains.

Conclusion

Advantages of the Approach

The proposed approach, which involves creating a semantic network based on the text and modifying it, has several significant advantages compared to directly modifying the text through an LLM request:

1. Control over Structure and Content. By modifying the semantic network, the analyst gains greater control over key concepts and the connections between them. This allows for preserving or emphasizing important aspects of the text, unlike in cases where an LLM may alter the text without clear consideration of priorities.
2. Ability for Deeper Modifications. The approach with semantic networks allows for flexible changes not only to the text but also to the conceptual structure. For instance, connections between certain ideas can be removed or new ones can be added. These changes can reflect in the text at a deeper level, which is hard to achieve through a direct request.
3. Process Transparency. Semantic networks enable transparent tracking of which concepts and connections are being modified, something that is more difficult when directly interacting with an LLM.
4. Preservation of Balance Between the Original Text and Adaptation The proposed approach allows for gradual text modifications through the semantic network, maintaining a balance between the original content and the modifications.
5. Automation and Use of Different Models. The analyst can experiment with different LLMs, generating multiple versions of the text based on a single modified semantic network.

Disadvantages of the Approach

1. Labor-intensive. Creating a semantic network and its subsequent modification require additional effort and time.
2. Need for additional knowledge. Working with semantic networks requires knowledge of conceptual relationships, analytical tools, and working with data structures.
3. Challenges in managing complex changes. Making significant changes to the semantic network may result in LLM generating text that significantly deviates from the original. Controlling these changes, especially in large texts or with numerous concepts, can become confusing.

4. Lack of immediate feedback. A direct request to LLM provides quick feedback, whereas creating and modifying a semantic network can take more time.

The proposed methodology for text generation through semantic network modification offers a novel approach to creating new meanings from existing content. The analogy to DNA and software modification underscores the potential and flexibility of this method. The paper provides a comprehensive framework, including mathematical formalism, application scenarios, and evaluation methods, for leveraging semantic networks and LLMs in text generation.

Thus, the approach of using semantic networks has significant advantages, particularly for a more controlled and transparent process of text modification. It provides more precise content management, which is important for adapting texts to specific goals or audiences. However, this method is more complex and time-consuming, requiring technical knowledge and time, which can be a drawback in situations where speed and simplicity are prioritized.

References

[1] Metje-Sprink J, Menz J, Modrzejewski D and Sprink T (2019) DNA-Free Genome Editing: Past, Present and Future. Front. Plant Sci. 9:1957. doi: 10.3389/fpls.2018.01957

[2] Francesco De Fazio, Conny Bakker, Bas Flipsen, Ruud Balkenende. The Disassembly Map: A new method to enhance design for product repairability. Journal of Cleaner Production. Volume 320, 20 October 2021, 128552. doi: 10.1016/j.jclepro.2021.128552

[3] Fu, Z., Lam, W., So, A. M.-C., & Shi, B. (2021). A Theoretical Analysis of the Repetition Problem in Text Generation. Proceedings of the AAAI Conference on Artificial Intelligence, 35(14), 12848-12856. doi: 10.1609/aaai.v35i14.17520

[4] Zitar, R. A., Al-Dmour, N., Nachouki, M., Hussain, H., & Alzboun, F. (2021). Hashing generation using recurrent neural networks for text documents. ICIC Express Letters, Part B: Applications, 12(3), 231-241. doi: 10.24507/icicelb.12.03.231

[5] Maria Johnsen. Large Language Models (LLMs). Independently published (June 15, 2024). - 451 p.ISBN: 979-8328536677

[3] Lande, D., & Strashnoy, L. (2023). GPT Semantic Networking: A Dream of the Semantic Web–The Time is Now. Kyiv, Engineering Ltd. ISBN - 168 p. ISBN: 978-966-2344-94-3