Generating 3D Models through Analyzing Natural Language Sentences with GPT API

Ye Jin Jin, Chae Yun Seo, and R. Young Chul Kim*
Software Engineering Laboratory, Graduate School, Hongik University
Sejong, South Korea
[e-mail: yejin_jin@g.hongik.ac.kr, chaeyun@hongik.ac.kr, bob@hongik.ac.kr]
*Corresponding author: R. Young Chul Kim

Abstract

As the 3D market expands, the cartoon industry is using 3D technology to produce 3D cartoons. Recently, AI image tools that generate background props or characters necessary for cartoons using only natural language input using AI technology have emerged. Background props can be used in various cartoons, but characters cannot be because of the author's intended style. To solve this, we propose an automatic mechanism for how to represent a facial emotion expression in a 3D object created. That is, we can make it possible for the emotional state of the main character in the cartoon to be reflected in the facial expression of the 3D character entity. The previous AI Text_to_3D (TT3D) tools generated various models for querying with each natural language, which cannot guarantee the accuracy of the results because the whole generation process is operated as a black box. To overcome these limitations, we try to perform a systematic natural language analysis and then visualize UML diagrams to copy with low visibility of natural language. We input natural language sentences and analyze them using the procedure through the GPT API. We apply software engineering to provide reusability, consistency, and traceability. With this, we expect to improve productivity and convenience in 3D cartoon development.

***Keywords*: GPT, Natural Language Processing, State Diagram, 3D Modeling, Cartoon Engineering**

1. Introduction

The 3D Modeling Market Size has been gradually increased recently[1]. The market growth is accelerating as 3D modeling is used in various industries such as architecture, construction, and games. In particular, the trend in the comics industry is to create webtoons through 3D modeling. 3D webtoons are characterized by realistic expressions compared to existing 2D methods, high quality, and fast work speed[2]. In general, 3D technology utilizing AI is mainly used for background design rather than character design. This is, because in comics, the same background or prop assets can be downloaded and used. However, it is difficult to apply this to characters. This is because a 3D model that reflects all changes in the character's facial expressions or behaviors depending on the scene is required. Since this process must reflect all the complex characteristics of the character, it is costly and time-consuming and requires expert-level skills.

Therefore, we systematically create 3D character entity from natural language. AI tools that create 3D models from natural language are already available, but they have limitations because they are difficult to satisfy detailed requirements. In addition, AI technology has problem with low maintenance and reusability because internal mechanisms cannot be confirmed. To solve this problem, we improved the existing mechanism and implemented it procedurally. We analyze the input natural language using GPT API, visualize it as UML, and then generate a 3D model containing the properties of the UML. This method enables tracking between natural language and models through UML as a middle step, and clearly understands the implementation process through a procedural approach.

The paper is organized as follows: Chapter 2 introduces the 3D modeling process and the existing proposed mechanism. Chapter 3 describes this research. Finally, Chapter 4 discusses the research and discusses future research directions.

2. Reconstruction of Volumetric Models

2.1 3D Modeling

3D character models are mainly used in the game industry, and recently, they are being applied to the cartoon industry. Character modeling involves many detailed steps, but broadly, it can be divided into six steps[3].

1) *Conceptual Design*: Determining the character's basic form and style. Establishing modeling plans by considering the character's personality and distinct features.

2) *Modeling*: Building the 3D form based on 2D designs. Visualizing the character's structure and expressing detailed features.

3) *Texturing and Rendering*: Visualizing the 3D model with textures and lighting. Adding color and texture to the character through texturing.

4) *Rigging*: Adding skeleton and joints to the 3D model. Configuring the character for natural movement.

5) *Animating*: Applying expressions and motions to the 3D model. Conveying the character's personality and emotions.

6) *Exporting*: Preparing the completed character for export in formats compatible with the target environment. Enabling integration with game engines or other 3D software.

In this study, we intend to perform steps 5 (Animating) and 6 (Exporting) using a 3D model that has already been completed from step 1 (Conceptual Design) to step 4 (Rigging). Through this, we study the process of adding emotional expressions and motions to an existing model and exporting it in a file format that can be utilized within a web environment.

2.2 Our Previous Mechanistic Study

In our previous study, we proposed a mechanism to generate 3D models from natural language[4]. Fig. 1 shows the process of the previous study.

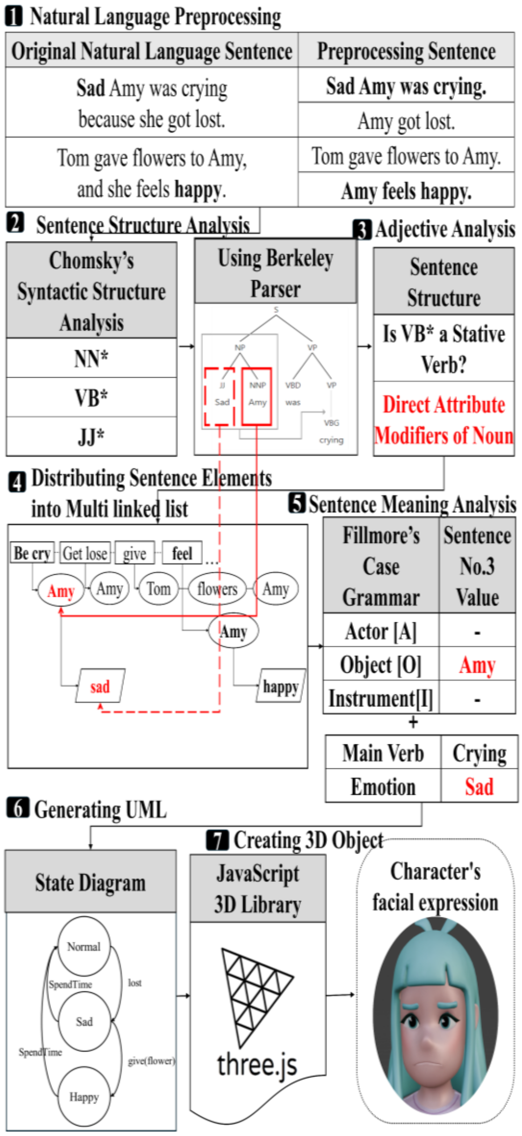

Fig. 1. Previous Study Process

We systematically analyzed natural language sentences, visualized them as UML, and generated 3D models based on the information to ensure traceability and consistency between natural language, UML diagrams, and 3D models. We presented a comparison between the proposed mechanism and commercial AI Text_to_3D(TT3D) tools. The AI tools could not generate identical characters or accurately express facial emotions. To solve this problem, we effectively analyzed the emotional state of the main character from natural language using state diagrams and expressed it as a 3D character entity. However, in previous studies, humans had to process this process manually. Therefore, we improved this and developed an automated tool through the LLM API.

3. Generating 3D Models from Text with GPT API

Our study improves the previously proposed mechanism from an implementation perspective. The research process is largely divided into three parts: 1) Natural language analysis, 2) UML diagram generation, and 3) 3D Character Entity generation. We apply LLM to the natural language analysis process to automate this process. The environment of the research that was performed is as shown in **Table 1**.

Table 1. Development Environment and Tools

Component	Version	Description
Node.js	20.14.0	JavaScript Runtime
Three.js	0.167.1	3D Library
Express	4.19.2	Web Framework
NLTK	3.5	NLP Library
OpenAI	1.15.2	GPT API

3.1 Natural Language Analysis

When a natural language sentence describing a cartoon scene is input, it is processed in three ways. First, the natural language sentence is preprocessed. Sentences consisting of complex and compound sentences are generally long, and the results of analyzing them are also extended. Therefore, complex sentences are converted into simple sentences. Simple sentences refer to sentences consisting of one subject and one verb.

Simple sentences are structurally analyzed using Chomsky's syntactic analysis theory[5]. Since it is essential to distinguish key morphemes in a sentence, the NLTK library is used to analyze the morphemes and part-of-speech tagging of the sentence. In addition, we analyze the type of structure the sentence exhibits. We analyze sentences into five structures, as shown in **Table 2**[6].

Table 2. Types of Sentences Structures

Sentence Type	Sentence Structure
Type 1	S+V
Type 2	S+V+C
Type 3	S+V+O
Type 4	S+V+I.O+D.O
Type 5	S+V+O+O.C

In **Table 2**, ‘S’ represents ‘Subject,’ ‘V’ for ‘Verb,’ ‘O’ for ‘Object,’ ‘I.O’ for ‘Indirect Object,’ ‘D.O’ for ‘Direct Object,’ ‘C’ for ‘Complement,’ and ‘O.C’ for ‘Objective Complement. After completing the structural analysis of all sentences, a JSON file is generated. This file contains the analyzed sentence, its order number, sentence structure type, morphemes, and parts-of-speech.

Based on the results, we analyze the sentence semantically. For semantic analysis, Fillmore's Sematic Roles theory is applied. Fillmore's theory analyzes the relationship between verbs and nouns and assigns appropriate roles to nouns[7]. The roles defined in the existing Fillmore theory have various interpretations[8]. Our study is to generate state diagrams through natural language in the cartoon domain. Therefore, we reduced the types of roles proposed by Fillmore and redefined the meanings as shown in **Table 3**.

Table 3. Redefined Fillmore’s Semantic Roles for State Diagram Generation

Semantic Roles	Definition
Actor	The instigator of the event/action.
Object	The entity which changes or existences is in consideration.
Instrument	The implement used in carrying out an event/action.

In order to assign a role to a noun, check the JSON file of the structural analysis result to see what kind of structure the sentence has and what the main verb is. As shown in **Table 3**, you can assign an appropriate role to the noun in the subject, object, etc. position depending on the structure and verb type. Once the semantic analysis is complete, it is returned as a JSON file. The JSON file consists of the analysis target sentence, the sentence order number, and the Main Verb, Emotional Adjective, and Fillmore Roles, which are the analysis results. **Fig. 2** shows the entire process of analyzing natural language. We applied the GPT API in the natural language analysis step and outputted the results in JSON format for each step.

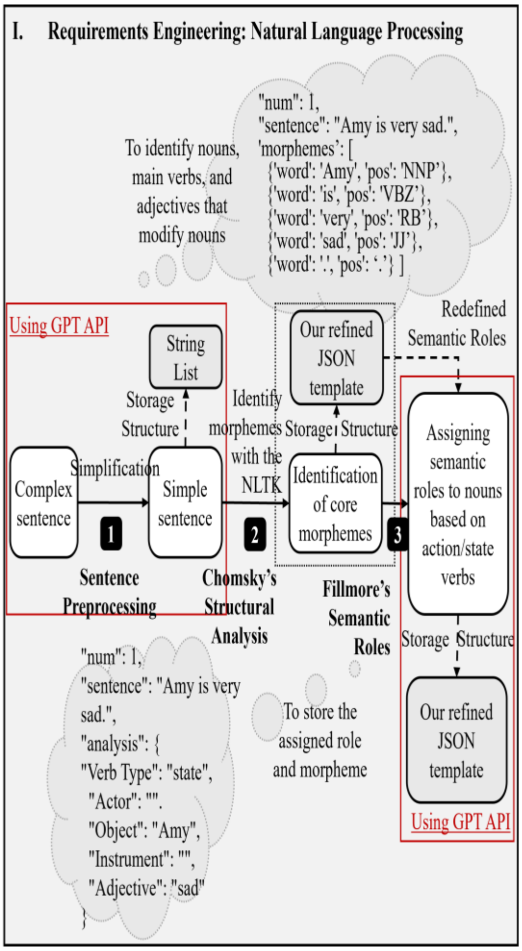

Fig. 2. Natural Language Analysis Process

3.2 UML State Diagram Generation

In order to analyze the emotional state changes of the main character, we create a UML State Diagram. Since we create diagrams using sentence analysis results, we map them.

In comics, since the character's emotions change depending on certain actions, 'Action Verb' is used as the 'Event' of the state diagram. The diagram is completed by writing 'Instrument' used for actions as 'Event's Parameter' and adjectives expressing emotions as 'State'. **Fig. 3** shows the process of generating a State Diagram from natural language.

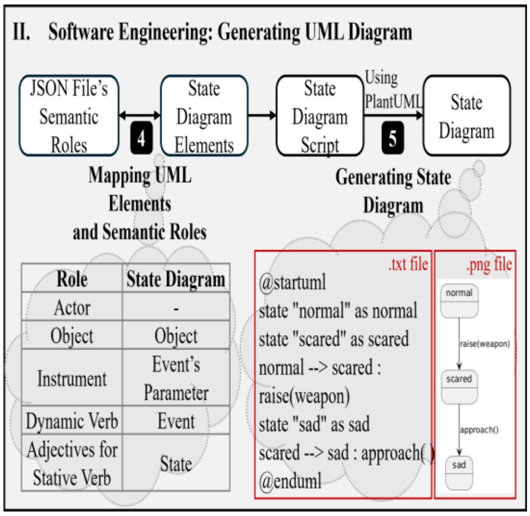

Fig. 3. Generating State Diagram from Semantic Roles

Using the Plant UML library, the diagram script (.txt) and diagram picture (.png) files are generated.

3.3 3D Model Generation in a 3D Environment

In order to express the emotional state of the 3D model from the state diagram, we load the generated diagram script file. We extract the diagram information from the file and apply it as the character's expression. There are five expressions: Normal, Joy, Sadness, Anger, and Fear, and they are indexed in order. The emotional expressions are stored as morph targets transformed from the original shape of the 3D model, and the degree corresponding to each emotional state can be adjusted from '0' to '1'. '0' means not applied at all, and '1' means fully applied. By adjusting this, various expressions corresponding to the emotional state can be implemented. **Fig. 4** shows this process.

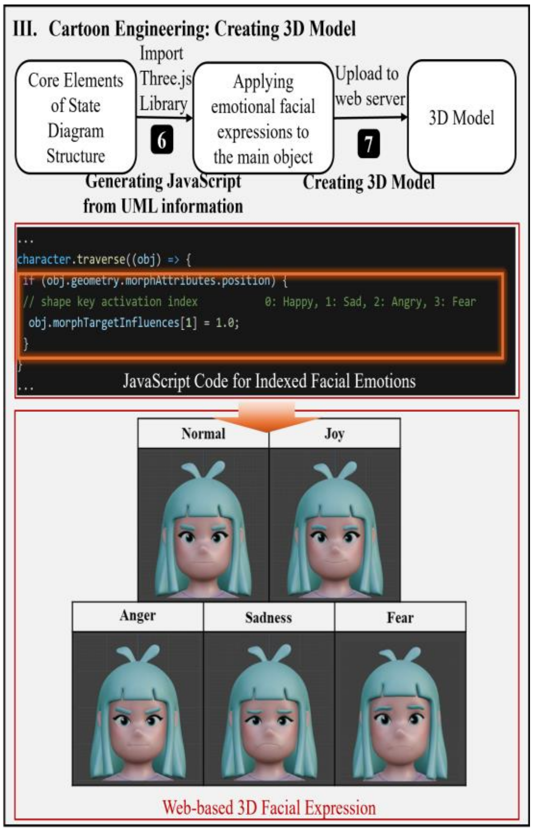

Fig. 4. Import 3D Models from the Web

4. Conclusions

This study improves the previously proposed mechanism for generating 3D models from natural language and attempts automation. We systematically analyze the natural language of comic scenes and convert them into 3D models. We analyze natural language analysis, which is difficult to process by rule-based method, using the GPT API. We generate state diagrams to make natural language sentences with low visibility easier to understand visually. In this process, we maintain consistency between natural language and 3D models by applying the information from the generated diagram to the expression of the 3D character entity. In addition, we automate some of the 3D modeling processes to provide that increases accessibility for non-experts in 3D modeling.

However, this study has several limitations. First, it requires a rigged 3D character entity. Second, we could only consider some sentence structures because natural language is diverse. Thus, our natural language analysis is limited. In future studies, we plan to improve natural language processing technology to process various sentence structures to overcome these limitations. In addition, we aim to completely automate the 3D model generation process so that 3D models can appear with only natural language input.

Acknowledgements

This research was supported by Korea Creative Content Agency (KOCCA) grant funded by the Ministry of Culture, Sports and Tourism (MCST) in 2024 (Project Name: Artificial Intelligence-based User Interactive Storytelling 3D Scene Authoring Technology Development, Project Number: RS-2023-0022791730782087050201) and National Research Foundation (NRF), Korea, under project BK21 Four.

References

[1] Fortune Business Insights. (2024, October 21). 3D Mapping & 3D Modeling Market Size, Share & Industry Analysis, 2024-2032. Retrieved from https://www.fortunebusinessinsights.com/3d-mapping-and-modeling-market-106003

[2] M.H Jeoung, "A Study on the 3D Computer Graphics Application in Webtoons," *Smart media journal*, Vol.4, No.3, pp.31-37, 2015.

[3] Kovanen, S., Functional Workflow in 3D Character Design, Karelia University of Applied Sciences, 2015.

[4] Y.J. Jin, C.Y. Seo, J.H. Kong, R. Y. C. Kim, "3D Object State Extraction Through Adjective Analysis from Informal Requirements Specs," *The Transactions of the Korea Information Processing Society (TKIPS)*, Vol.13, No.10, pp.529-536.

[5] N. Chomsky, Syntactic structures, USA: Mouton de Gruyter, 2002.

[6] Betti, M. J., Sentence Patterns in English, University of Thi-Qar. 2021.

[7] C. J. Fillmore, The Case for Case, New York: *HR&W*, 1968.

[8] B. Park, and R. Y. Kim, "Effort estimation approach through extracting use cases via informal requirement specifications," *Applied Sciences*, vol.10, No.3044, pp.1-15, 2020.