A Proposal to Create a Corpus of Patent Claim Analysis Results Expressed in Mermaid Notation using Generative AI

Akihiro Shinmori
Toyama University of International Studies
65-1 Higashi Kuromaki, Toyama 930-1292 JAPAN
shinmori@tuins.ac.jp

Abstract

To address the issue of poor readability of patent claims which describes the scope of patent rights, and the "Claims" section, which lists one or more claims, I performed citation relationship analysis and structural analysis of Japanese patent claims using generative AI. I used the Gemini API as the generative AI and employed a few-shot prompting method using examples from my past papers and examples created from analysis failures. The analysis results were generated in Mermaid notation and registered on GitHub, making them visible in a browser. Based on this experiment, I propose to create a corpus of patent claim analysis results on GitHub public repository. It will be possible for everyone to contribute to create, enhance, and share patent claim analysis results. By doing so, patent documents will become more accessible to engineers and the public who are not intellectual property specialists, and it will lead to technological advancement.

Keywords: patent claim analysis, Mermaid notation, GitHub

1. Introduction

Patents play a crucial role as intellectual property in every engineering. Patent documents include the "Claims" section which list one or more patent claims. While this section is the most important part of the document, its poor readability has been pointed out, and solutions using natural language processing techniques have been proposed (Casolla and Lavelli, 2022).

I addressed this issue approximately twenty years ago by conducting research on the structural and citation relationship analysis of patent claims (Shinmori et al., 2023; Shinmori, Ooya, and Tanigawa, 2008). However, the methods employed at that time predated the advent of generative AI, utilizing techniques such as morphological analysis, context-free grammars, and rhetorical structure analysis. The tool used to visualize the analysis results was RSTTool (O'nnel, 1997), a dedicated application for displaying rhetorical structure analysis outputs.

In the field of natural language processing, methods based on deep learning subsequently became the mainstream, and in recent years, generative AI has been utilized to its full potential. Meanwhile, within software engineering, the Mermaid notation[1], a text-based diagramming syntax and its associated graphing tools have been widely used.

In this study, citation relationship analysis and structural analysis were performed on twenty patent claims randomly selected from published patent gazettes released in 2024.

I used Gemini 2.5 flash[2] for both analysis by using few-shot prompts[3] created using examples of my previous two papers. The outputs were generated as flowchart-style graphs using Mermaid notation. I uploaded the graphs to GitHub because GitHub supports Mermaid notation in the markdown files (Woodward and Biagianti, 2022).

2. Analysis of "Claims" by Using Generative AI

2.1 Citation Relationship Analysis of Claims

Figures 1 and 2 show examples of the results of patent claim citation relationship analysis displayed in the Chrome browser. In both figures, the analysis results were registered on GitHub and displayed as graphs by writing Mermaid notation in the README.md file. In both figures, the fact that "claim n cites claim m" is represented by drawing an arrow from claim n to claim m. In addition, the URI (Uniform Resource Indicator) for the document on Google Patents[4] is written at the beginning so that the document can be accessed immediately.

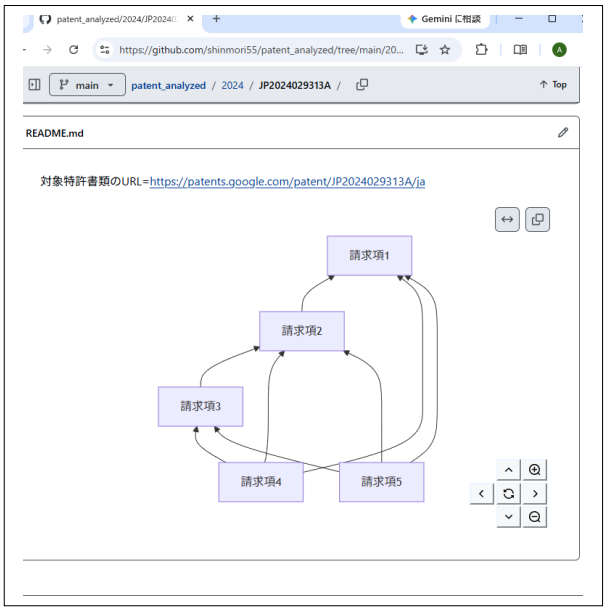

Figure 1: Results of citation relationship analysis of patent claims in JP 2024-29313 (shown in Chrome browser)

[1] https://en.wikipedia.org/wiki/Mermaid_(software)
[2] https://gemini.google.com/app
[3] https://cloud.google.com/discover/what-is-prompt-engineering?
[4] https://patents.google.com/

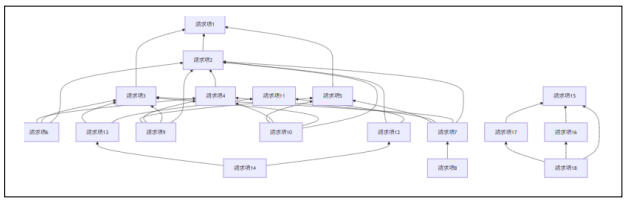

Figure 2: Results of citation relationship analysis of patent claims in JP 2024-284 (extracted from the graph display portion shown in the Chrome browser)

Figure 1 shows a patent with five claims. Only claim 1 is an independent claim and the other four are cited claims. While there are instances of multiple citations, the citation relationships are simple.

Figure 2 shows a patent with eighteen claims, where claims 1 and 15 are independent claims. The citation relationships of the remaining sixteen claims are complex, including "multiple-multiple citations" where the source of a multiple citation itself is a multiple citation. Manually writing such patent claim citation relationship diagrams is a time-consuming task; therefore, the ability to automatically generate them using a generation AI is highly significant.

Visual verification of the citation relationship analysis of twenty patent claims revealed that all were correctly analyzed. In other words, the correct analysis results could be output without modifying the prompt created using the example from the 2008 paper (Shinmori, Ooya, and Tanigawa, 2008).

2.2 Structure Analysis of Claim

Examples of structural analysis results for patent claims are shown in Figures 3, 4, and 5.

Figure 3 shows an independent claim. The phrase "であって、" and its modifying part, form the first part, followed by the second part. The second part is further divided into two sections, with the modifying parts clearly indicated. The first part of the second part also clearly indicates that the processing device consists of two parts.

Figure 4 also shows an independent claim. A parallel structure consisting of five parts is recognized, and the first part further recognizes two parts: "配線部材" and "線状伝送部材."

Figure 5 shows a cited claim. The phrase "請求項2に記載の" is merged with "配線部材," clearly indicating that it is a cited claim.

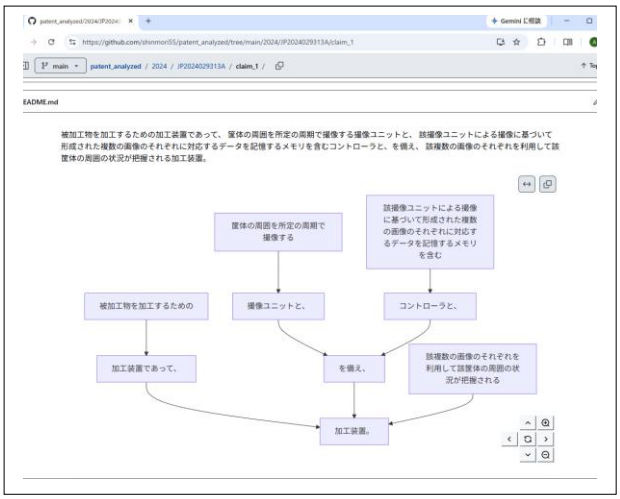

Figure 3: Structural analysis results of the first claim of Japanese Patent Publication No. 2024-29313 (shown in Chrome browser)

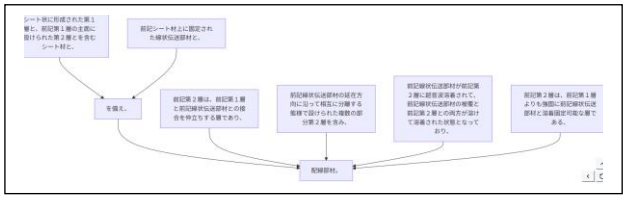

Figure 4: Structural analysis results of the first claim of Japanese Patent Publication No. 2024-1238 (Extracted only the graph portion from Chrome browser.)

Figure 5: Structural analysis results of claim 3 of Japanese Patent Publication No. 2024-1238 (shown in Chrome browser)

3. Creating a Corpus of Patent Claim Analysis Results

The results of the citation relationship analysis and structural analysis should be shared with relevant parties and widely used, so the prompts and analysis results used in this research were stored and made public in a GitHub public repository[5]. The structure of the repository is shown in Figure 6. It is hoped that in the future, the results of citation relationship analysis and structural analysis of many publications will be stored in this manner, creating a "publication analysis result corpus."

Git and GitHub have the concept of "commit messages," which allow you to write a description when registering (committing) a newly created file. Since I thought that the model and prompt used to obtain the analysis results should be clearly indicated, I decided to use commit messages to describe "which model and prompt were used for the analysis" in the analysis results. Because the prompts are modified several times through trial and error, they are subject to GitHub version control, and a permalink[6], which is an immutable link to a specific version of the file, is included in the commit message. This concept is shown in Figure 7.

```
patent_analyzed /
├── README.md
├── prompt_citation_analyzer.txt  # Text file containing prompt for
│                                 # citation relationship analysis
├── prompt_claim_analyzer.txt     # Text file containing prompt for
│                                 # citation relationship analysis
├── 2024/
│   ├── JP2024000284A         # Subdirectory for JP24000284A
│   │   ├── README.md         # Result of citation relationship analysis
│   │   ├ claim_1
│   │   ├── README.md         # Result of structure analysis of claim 1
│   │   ├── claim_1.txt       # Text file of claim 1
│   │   ├ claim_2
│   │   …
│   ├── JP2024001238A         # Subdirectory for JP24001238A
│   …
```

Figure 6: Sample repository structure of GitHub as a corpus of Patent Claim Analysis Results

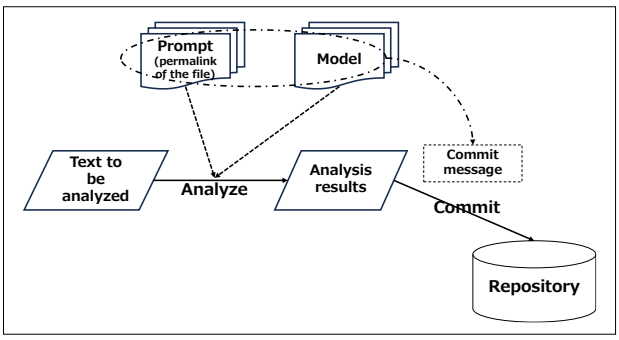

Figure 7: Conceptual diagram of using commit message indicating prompt and model used.

4. Conclusion

The motivation for this research was to investigate how well the patent claim analysis performed approximately twenty years ago could be done using generative AI. The current conclusion is that generative AI can obtain the same level of results.

It has already been reported in the natural language community that generative AI can perform tasks equivalent to or even exceeding those performed by conventional language processing. In that sense, the analytical aspects of this research are a reaffirmation of these reports.

The contribution of this research lies, firstly, in analyzing not only the structure of patent claims but also citation relationships and displaying the results graphically using Mermaid notation. Furthermore, with the aim of creating a "publication analysis result corpus," the analysis results were published on GitHub, and the graphical display was also utilized using GitHub's features.

5. Acknowledgements

This research was supported by a research grant from the Toyama First Bank Scholarship Foundation. Discussions held within the Document Creation Subcommittee of the Industrial Japanese Research Group (chaired by Dr. Koichi Hashida) provided valuable insights during this research. I wish to express my sincere gratitude.

6. Bibliographical References

Casolaa, S. and Lavelli, A. (2022). Summarization, Simplification, and Generation: The Case of Patents, https://arxiv.org/pdf/2104.14860.

OD'nnel, M. (1997). RST-Tool: An RST Analysis Tool. *Proceedings of the 6th European Workshop on Natural Language Generation*.

Sheremeteyeva, S. (2014). Automatic Text Simplification For Handling Intellectual Property (The Case of Multiple Patent Claims), Proceedings of the workshop on automatic text simplification-methods and applications in the multilingual society (ATS-MA 2014).

Shinmori, A. Okumura, M. Marukawa, Y. and Iwayama, M. (2003). Patent Claim Processing for Readability - Structure Analysis and Term Explanation-. ACL-2003 Workshop on Patent Corpus Processing, Association for Computational Linguistics.

Shinmori, A., Ooya, Y., Tanigawa, H. (2008). Detecting and Rewriting Multiple-multiple Dependent Form Claims, Journal of Information Processing, Information Processing Society of Japan. (in Japanese).

Woodward, M. and Biagianti, A. (2022). Include diagrams in your Markdown files with Mermaid. https://github.blog/developer-skills/github/include-diagrams-markdown-files-mermaid/

[5] https://github.com/shinmori55/patent_analyzed

[6] https://docs.github.com/en/repositories/working-with-files/using-files/getting-permanent-links-to-files

Appendix

A1. Prompt used for citation relationship analysis

Prompt used for citation relationship analysis is shown on the following GitHub page:

https://github.com/shinmori55/patent_analyzed/blob/7f197e5b020cf7e1daff419e733f680bd07fdd96/prompt_citation_analyzer.txt

A2. Prompt used for structure analysis

Prompt used for structure analysis is shown on the following GitHub page:

https://github.com/shinmori55/patent_analyzed/blob/7f197e5b020cf7e1daff419e733f680bd07fdd96/prompt_claim_analyzer.txt

A3. Mermaid Notation of Figure 1

```
flowchart BT
A[請求項 1]
B[請求項 2]
C[請求項 3]
D[請求項 4]
E[請求項 5]
B --> A
C --> B
D --> A
D --> B
D --> C
E --> A
E --> B
E --> C
```

A4. Mermaid Notation of Figure 3

```
flowchart TD
A[被加工物を加工するための]
B[加工装置であって、]
C[筐体の周囲を所定の周期で撮像する]
D[撮像ユニットと、]
E[該撮像ユニットによる撮像に基づいて形成された複数の画像のそれぞれに対応するデータを記憶するメモリを含む]
F[コントローラと、]
G[を備え、]
H[該複数の画像のそれぞれを利用して該筐体の周囲の状況が把握される]
I[加工装置。]
A --> B
B --> I
C --> D
E --> F
D --> G
F --> G
G --> I
H --> I
```