# Can Claim Analysis Contribute toward Patent Map Generation?

Akihiro SHINMORI
INTEC Web and Genome Informatics Corp.
3-23 Shimoshin-machi
Toyama 930-0804, JAPAN
shinmori_akihiro@webgen.co.jp

Manabu OKUMURA
Precision and Intelligence Laboratory
Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku,
Yokohama 226-8503, JAPAN
oku@pi.titech.ac.jp

Yuzo MARUKAWA
Core Research for Evolutional
Science and Technology,
Japan Science and Technology Agency
maru@nii.jp

Makoto IWAYAMA
Precision and Intelligence Laboratory
Tokyo Institute of Technology, and
Hitachi, Ltd.
iwayama@pi.titech.ac.jp

## Abstract

*An effort for the patent map generation task in NT-CIR4 is reported. The attempt to utilize the research of claim analysis was failed and an tentative approach was taken. The results were evaluated by human experts and it suggests the need for more research in this field.*
**Keywords:** *claim analysis, alignment, term extraction*

## 1 Introduction

In participating in the patent map generation task of NTCIR4 [2], we tried to take advantage of our previous research on claim analysis [3, 4]. But, in the course of doing the task, we found that with the current level, the result of claim analysis cannot be used effectively for the task.

In this paper, we first present our research on claim analysis briefly. Next, we present what we actually did for the patent map generation task and analyze the result. Finally, we present some discussion about the result.

## 2 Research on Claim Analysis

We have been researching on claim analysis to improve readability. What we have done so far is to analyze claim structure by using cue phrase [3, 4] and to align each core element acquired by the analysis with each sentence in detailed description.

For example, for a claim in Figure 1, the structure can be analyzed as in Figure 2 and two of the core el-

ement of the claim can be aligned with the sentence in Figure 3. By the alignment, the following two paraphrases can be obtained.

**Paraphrase 1**

- (hitsuyou sei wo kettei suru) [decide the necessity]

- (hitsuyou to shite iruka wo ketteishi) [decides which nozzles need to be maintained]

**Paraphrase 2**

- (ink funshutsu nozzle kara purge you ink eki wo funshutsu suru) [jet purging ink drops from ink jetting nozzles]

- (ink wo purge suru) [purges ink]

At first, we thought we could extend the above research to be effectively applied toward the patent map generation. But, when we analyzed the task in detail, we decide that it was impossible to apply directly the alignment research with the current level. Therefore we took a tentative approach described in the next section.

## 3 Our Tentative Approach

For the patent map generation task, we used the full-text of patent specifications that have the following structure.
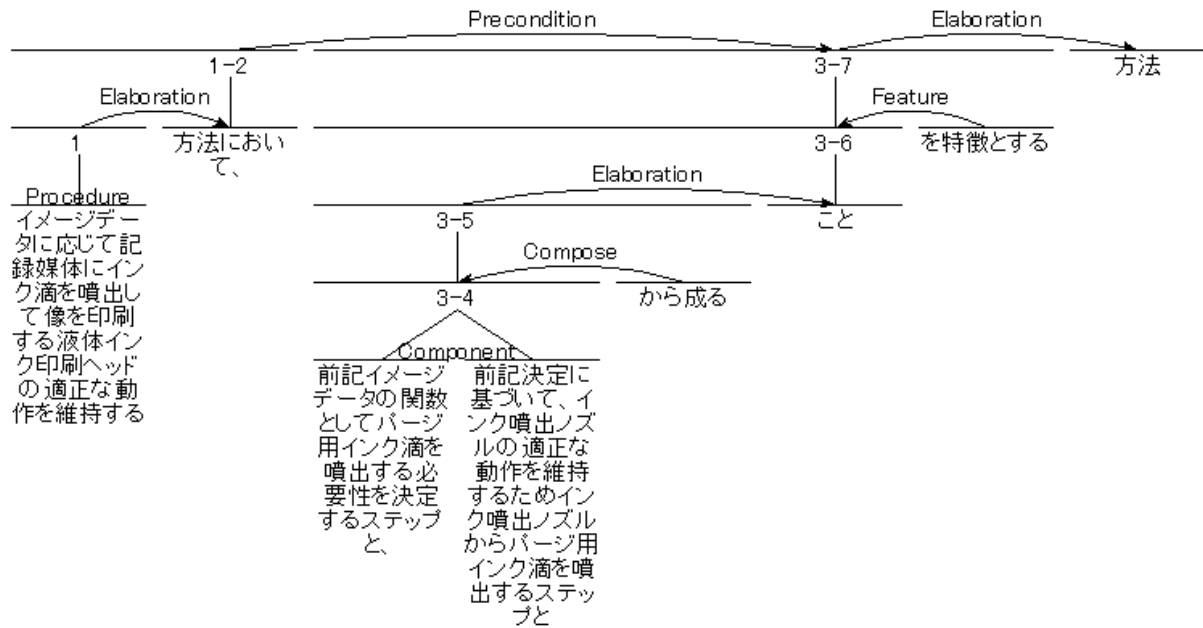
Precondition    Elaboration

1-2    3-7    方法

Elaboration    Feature

1    方法におい    3-6    を特徴とする
     て、

Procedure    Elaboration

イメージデー    3-5    こと
タに応じて記
録媒体にイン    Compose
ク滴を噴出し
て像を印刷    3-4    から成る
する液体イン
ク印刷ヘッド    Component
の適正な動
作を維持する    前記イメージ    前記決定に
     データの関数    基づいて、イ
     としてパージ    ンク噴出ノズ
     用インク滴を    ルの適正な
     噴出する必    動作を維持
     要性を決定    するためイン
     するステップ    ク噴出ノズル
     と、    からパージ用
          インク滴を噴
          出するステッ
          プと

**Figure 2. A result of structure analysis of the claim in Figure 1**

(An method to maintain appropriate motion of liquid ink printing heads which jet ink drops to printing materials based on image data, comprising:
(a) a step which decides the necessity of jetting ink based on said image data;
(b) a step which jet purging ink drops from ink jetting nozzles to maintain appropriate motion of ink jetting nozzles.)

**Figure 1. A sample Japanese claim extracted from a patent (publication number=10-6528)**

(Considering the above problems, this invention provides a device and a method which decides which nozzles need to be maintained and which purges ink only from those nozzles, instead of purging from the array of all nozzles. )

**Figure 3. The aligned sentence in the detailed description for the claim in Figure 1**

- Invention Title
- Summary
- Claims
- Detailed Description
    - Field of the invention
    - Prior art
    - Means of solving the problems
    - Embodiments of the invention
    - Effects of the invention
- Brief Explanation of Drawings

The first thing we did was to decide the phrases for the x and the y axes. To do so, we first looked

through the "prior art" and the "means of solving the problems" to pick up core nouns. Next, we ran a program which can pick up compound nouns including the above nouns. The candidate nouns and the compound nouns are selected as keywords which are shown in Table1. Finally, we ran a program to search patent specifications to find the occurrence of the keywords in the summary and the claims.

In the following, each topic is described:

**Topic 007** The topic 007 consists of a collection of patents about the "gasoline direct-injection engine". The x axis was requested to set the "keywords which express the concave" and the y axis was requested to set the "keywords (notations/spells) which express the piston top face."

**Topic 008** The topic 008 consists of a collection of patents about "hair care cosmetic products". The x axis was requested to set the "forms of the products (liquid etc.)" and the y axis was requested to set the "date of publication".

**Topic 010** The topic 010 consists of a collection of patents about "functional carpet". The x axis was requested to set the "problems to be solved (chemical substances" and the y axis was requested to set the "solutions (materials)".

**Topic 012** The topic 012 consists of a collection of patents about "blue light-emitting diode". The x axis was requested to set the "problems to be solved" and the y axis was requested to set the "solutions".

**Topic 024** The topic 024 consists of a collection of patents about "solid high-polymer-type fuel cell". The x axis was requested to set the "problems to be solved" and the y axis was requested to set the "solutions".

**Topic 025** The topic 025 consists of a collection of patents about "ultra hydrophilization of plastic surfaces". The x axis was requested to set the "problems to be solved" and the y axis was requested to set the "solutions".

## 4 Result

As a feasibility study, human experts evaluated the result. In addition, model answers created by human experts were given.

While some experts evaluated positively about the result, others evaluated negatively saying that the analysis was too coarse to be useful.

In the following, the evaluation for each topic is summarized:

**Topic 007** The expert evaluated the result for topic 007 positively saying, "The keywords for the x and the y axes were correctly chosen and the map can be used as the starting point to investigate technology trend." But, it was pointed out that some of the keywords are incorrectly extracted from the "prior arts" and the "embodiments". In the model answer, three items were chosen for the x axis as "　　　　　^", "　　　", and "　^" [1]. The result did not contain two of them ("　　　^" and "　　　") and included two items ("　　　" and "　　　") which were not chosen in the model answer.

**Topic 008** The evaluator said that "　　　　　　" [shampoo] and "　　　　　　　　" [conditioner] should not be regarded as "forms of the products" because they represent "applications". In the model answer, six items were chosen for the x axis as "　　　(　　　　　　　　　　)" [general forms in which the forms are irrelevant], "　　　　　　　" [forms in immersion], "　　　" [forms in solid], "　　　　" [forms in powder], "　　　　" [forms in liquid], and "　　　　" [other forms].

**Topic 010** The evaluation was not given for the topic 010. In the model answer, ten items were chosen for the x axis as "　　　　　　　" [acetaldehyde], "　　　　　　　　" [formaldehyde], "　　　　　" [ammonia], "　　　　　　　　" [trimethylamine], "　　　　　" [hydrogen sulfide], "　　　　　　　" [methylkaptin], "　　　" [ethylkaptin], "　　　" [acetic acid], "　　　" [toluene], and "　　　　" [pyridine]. The result contained three of them. In the model answer, ten items were chosen for the y axis as "　　　(　　　　　　　)", "　　　　　　(　　　　)", and "　　　(　　　　　　)". The result does not contain such high-level descriptions.

**Topic 012** The evaluator said that the analysis was too coarse to be useful. For the 121 lines the system generated, 18 (14.9%) were evaluated as correct, 49 (40.5%) were evaluated as partially correct, and 29 (24.0%) were evaluated as incorrect.

**Topic 024** The evaluator said that the analysis was too coarse to be useful. For the 149 lines the system generated, 10 (6.7%) were evaluated as correct, 67 (45.0%) were evaluated as partially correct, and 22 (14.8%) were evaluated as incorrect.

**Topic 025** The evaluator said that the analysis was too coarse to be useful. He also said that as the tech-

---

[1] "^" means any characters.

## Table 1. Keywords of the x axis for each topic

| Topic | Axis | Keywords |
|---|---|---|
| 007 | x | " "[concave type] , " "[concave type], " "[concave groove] , " "[concave place], " "[concave setting], " "[concave part], " "[concave cavity], " "[concave curved surface], " "[concave type], " "[concave part], " "[concave spce], " "[concave shape], " "[concave shape], " "[concave area], " "[concave shaping part], " "[concave combustion room], " "[concave], " " [concave place]. |
| 007 | y | " "[top of piston], " "[top of piston], " "[top of piston], " "[top of piston]," "[top of piston], " "[top of piston], " "[top of piston], " "[top of piston], " "[top of piston], " "[top of piston] |
| 008 | x | " "[foam material], " "[fixation material], " "[humidification material], " "[cleaner]," "[hair remover]," "[deodorant], " "[softener], " "[make-up material], " "[mousse], " " [cologne]," "[scent bag], " "[shampoo]," "[hair protectant]," " [hair conditioner], " "[color make-up], " "[detergent]," " [hair-care material], " "[hair-set material], " "[hair make-up], " "[permanent wave material], " "[immersion], " "[hair spray], " "[hair shampoo], " "[shampoo-type protectant], " "[sunscreen], " "[moisturizer], " "[hair setting material], " "[cuticle coat], " "[hair conditioner] |
| 008 | y | (The "date of publication" field extracted from the patent specifications) |
| 010 | x | " "[acetaldehyde], " " [formaldehyde], " "[basic effluvium component], " " [neutral effluvium component], " "[acidic effluvium component], " " [ammonia]," "[amine type], " "[hydrogen sulfide]," "[melkaptin type], " "[carbonyl group], " "[acidic chemical compound], " "[basic chemical compound] |
| 010 | y | " "[anatase titanium oxide], " "[odor eliminating fiber], " "[acrylic-type synthetic fabric], " "[acrylic nitrile], " "[adipine acid dihydradido], " "[trichloro calban type chemical compound], " "[hydradin derivative], " "[odor eliminating inorganic material], " "[hydradido chemical compound], " "[polyamine chemical compound], " "[polyester type fiber material], " "[polyolefin crimp fiber], " "[benzoic acid type material], " "[functional fiber], " "[acidic zirconium], " "[polymer with acid anhydride base], " "[plastic composition], " "[odor eliminator]," "[odor eliminating compound], " "[odor eliminating fiber], " "[hydration acidic zirconium], " "[chalco-compound]," "[silicon dioxide], " "[aromatic polyamine] , " "[inorganic chemical compound] |
| 012 | x | " "[luminescence intensity], " "[luminescence efficiency], " "[process yield], " "[white luminescence]," "[cost], " "[light intencity], " "[light output], " "[low power], " "[light pattern] |
| 012 | y | " "[nitride semiconductor]," "[indium], " "[nitride gallium], " "[zinc oxide], " "[double-hetero structure] |
| 024 | x | " "[cost]," "[efficiency]," "[durability]," " [stability] |
| 024 | y | " "[solid giant molecule]," "[solid ionic conductor] , " "[solid electrolyte], " "[solid giant molecule electrolyte], " "[giant molecule electrolyte], " "[methanol]," "[hydrophilia material], " "[graphite]," "[fuluorine] |
| 025 | x | " "[fog]," "[fog]," "[stain]," "[water drop], " "[defogging]," "[anti-moisture], " "[anti-stain]," "[flipping phenomenon], " "[hardness]," "[transparence], " "[anti-wearing-out] |
| 025 | y | " "[titanium oxide]," "[oxide compound]," "[metal oxide], " "[silicon oxide]," "[molybdic oxide]," "[tungsten oxcide], " "[water-repellent fluorocarbon resin], " "[amorphous silica], " "[light catalyst] |

nology for the topic 025 was simple, the result given by the system was not so bad. For the 380 lines the system generated, 64 (16.8%) were evaluated as correct, and 35 (9.2%) were evaluated as partially correct.

## 5 Discussion

By participating the patent map generation task, we learned how difficult it is to automatically generate a map that can satisfy human experts.

The comment given for the topic 008 suggests the necessity of ontology. The evaluator for the topic 012 and 025 suggests that "multi-level analysis" for the solution was required. The evaluator for the topic 008 said that he weighs the keywords depending on the nature of applicant (whether they manufacture cosmetic products or materials).

## 6 Concluding Remarks

Automatic patent map generation is a difficult task and a lot of research effort is required for the system to be practically useful.

For the question we raised in the title, we have to answer "no" at the current level of claim analysis. But, as the claims are the most important in patent specifications [1], we believe the result of more sophisticated claim analysis should be incorporated into the effort of patent map generation.

## Acknowledgement

## References

[1] L. B. Burgunder. *Legal Aspects of Managing Technology*. South Western, 1995.

[2] A. Fujii, M. Iwayama, and N. Kando. Overview of patent retrieval task at ntcir-4. In *NTCIR4 Working Notes*, 2004.

[3] A. Shinmori, M. Okumura, Y. Marukawa, and M. Iwayama. Rhetorical structure analysis of japanese patent claims using cue phrases. In *Proceedings of the Third NTCIR Workshop on research in information Retrieval, Automatic Text Summarization and Question Answering*. NII (National Institute of Informatics), 2002.

[4] A. Shinmori, M. Okumura, Y. Marukawa, and M. Iwayama. Patent claim processing for readability - structure analysis and term explanation. In *Proceedings of the Workshop on Patent Corpus Processing (ACL 2003 Post-Conference Workshop)*, pages 56–65, 2003.