

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

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

Manabu Okumura
Precision and Intelligence Laboratory
Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku,
Yokohama 226-8503, JAPAN
oku@pi.titech.ac.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 NTCIR4 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

- 必要性を決定する
(hitsuyou sei wo kettei suru) [decide the necessity]
- 必要としているかを決定し、
(hitsuyou to shite iruka wo ketteishi) [decides which nozzles need to be maintained]

Paraphrase 2

- インク噴出ノズルからパージ用インク滴を噴出する
(ink funshutsu nozzle kara purge you ink eki wo funshutsu suru) [jet purging ink drops from ink jetting nozzles]
- インクをパージする
(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.

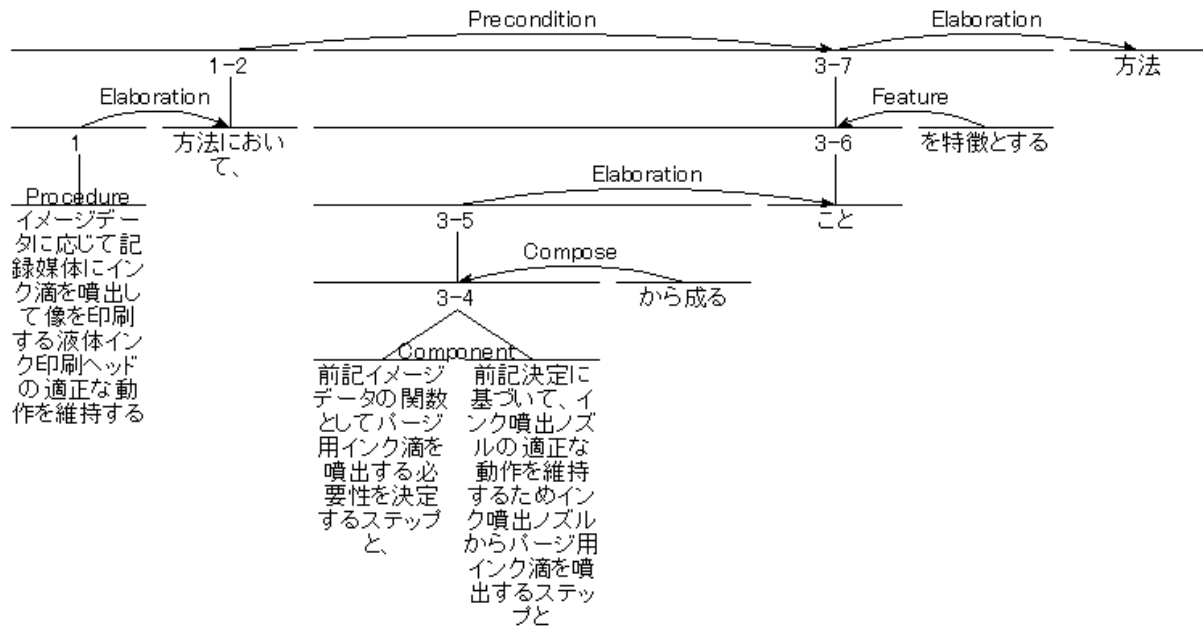


Figure 2. A result of structure analysis of the claim in Figure 1

イメージデータに応じて記録媒体にインク滴を噴出して像を印刷する液体インク印刷ヘッドの適正な動作を維持する方法において、前記イメージデータの関数としてパージ用インク滴を噴出する 必要性を決定するステップと、 前記決定に基づいて、インク噴出ノズルの適正な動作を維持するため インク噴出ノズルからパージ用インク滴を噴出するステップと から成ることを特徴とする方法。

(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)

これらの問題を考慮して、本発明は、全ノズルアレイからインクをパージする代わりに、印刷ヘッド内のどのノズルがメンテナンスを 必要としているかを決定し、それらのノズルだけから選択的にインクをパージする装置および方法を提供する。

(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 Table 1. 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 “キャビテ[^]”, “傾斜”, and “凹[^]”¹. The result did not contain two of them (“キャビテ[^]” and “傾斜”) and included two items (“窪み” and “窪部”) which were not chosen in the model answer.

Topic 008 The evaluator said that “シャンプー” [shampoo] and “コンディショナー” [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 “一般 (形態が重要でないもの)” [general forms in which the forms are irrelevant], “エマルジョン系” [forms in immersion], “固形状” [forms in solid], “粉末状” [forms in powder], “液体状” [forms in liquid], and “その他” [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 “アセトアルデヒド” [acetaldehyde], “ホルムアルデヒド” [formaldehyde], “アンモニア” [ammonia], “トリメチルアミン” [trimethylamine], “硫化水素” [hydrogen sulfide], “メチルカプタン” [methylkaptin], “エチルカプタン” [ethylkaptin], “酢酸” [acetic acid], “トルエン” [toluene], and “ピリジン” [pyridine]. The result contained three of them. In the model answer, ten items were chosen for the y axis as “無機化合物 (+ 高分子化合物混合物、有機化合物混合物、金属混合体、但し、有機金属錯体は除く)”, “高分子化合物 (無機化合物との混合を除く)”, and “その他 (有機低分子化合物、有機金属錯体ほか)”. 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	<p>“凹型”[concave type], “凹形”[concave type], “凹溝”[concave groove], “凹所”[concave place], “凹設”[concave setting], “凹部”[concave part], “凹窩”[concave cavity], “凹曲面”[concave curved surface], “凹形状”[concave type], “凹状部”[concave part], “凹部空間”[concave space], “凹部形状”[concave shape], “凹部形成”[concave shape], “凹部領域”[concave area], “凹部形成部”[concave shaping part], “凹状燃焼室”[concave combustion room], “窪み”[concave], “窪部” [concave place].</p>
007	y	<p>“ピストン上面”[top of piston], “ピストン上部”[top of piston], “ピストン冠面”[top of piston], “ピストン冠部”[top of piston], “ピストン頂面”[top of piston], “ピストン頂部”[top of piston], “ピストンの頂面”[top of piston], “ピストン頂面部”[top of piston], “ピストン最上部”[top of piston], “ピストン上面部”[top of piston]</p>
008	x	<p>“泡剤”[foam material], “固定剤”[fixation material], “加湿剤”[humidification material], “洗淨剤”[cleaner], “脱毛剤”[hair remover], “防臭剤”[deodorant], “柔軟剤”[softener], “化粧料”[make-up material], “ムース”[mousse], “コロン水” [cologne], “におい袋”[scent bag], “シャンプー”[shampoo], “髪用保護剤”[hair protectant], “毛髪処理剤” [hair conditioner], “着色化粧品”[color make-up], “界面活性剤”[detergent], “ヘアケア剤” [hair-care material], “ヘアセット剤”[hair-set material], “毛髪用化粧品”[hair make-up], “パーマメント”[permanent wave material], “エマルジョン”[emulsion], “ヘアースプレー”[hair spray], “ヘアシャンプー”[hair shampoo], “洗淨型髪用保護剤”[shampoo-type protectant], “サンスクリーン剤”[sunscreen], “モイスチャライザー”[moisturizer], “毛髪セッティング剤”[hair setting material], “キューティクルコート”[cuticle coat], “ヘアコンディショナー”[hair conditioner]</p>
008	y	<p>(The “date of publication” field extracted from the patent specifications)</p>
010	x	<p>“アセトアルデヒド”[acetaldehyde], “ホルムアルデヒド” [formaldehyde], “塩基性臭気成分”[basic effluvium component], “中性臭気成分” [neutral effluvium component], “酸性臭気成分”[acidic effluvium component], “アンモニア” [ammonia], “アミン類”[amine type], “硫化水素”[hydrogen sulfide], “メルカプタン類”[melkaptin type], “カルボニル基”[carbonyl group], “酸性化合物”[acidic chemical compound], “塩基性化合物”[basic chemical compound]</p>
010	y	<p>“アナターゼ型酸化チタン”[anatase titanium oxide], “消臭繊維”[odor eliminating fiber], “アクリル系合成繊維”[acrylic-type synthetic fabric], “アクリロニトリル”[acrylic nitrile], “アジピン酸ジヒドラジド”[adipine acid dihydradido], “トリクロロカルバン系化合物”[trichloro calban type chemical compound], “ヒドラジン誘導体”[hydradin derivative], “消臭性無機物質”[odor eliminating inorganic material], “ヒドラジド化合物”[hydradido chemical compound], “ポリアミン化合物”[polyamine chemical compound], “ポリエステル系繊維材料”[polyester type fiber material], “ポリオレフィン捲縮繊維”[polyolefin crimp fiber], “安息香酸系加工物”[benzoic acid type material], “機能性繊維”[functional fiber], “酸化ジルコニウム”[acidic zirconium], “酸無水物基を有するポリマー”[polymer with acid anhydride base], “樹脂組成物”[plastic composition], “消臭剤”[odor eliminator], “消臭性樹脂組成物”[odor eliminating compound], “消臭性繊維”[odor eliminating fiber], “水和酸化ジルコニウム”[hydration acidic zirconium], “銅系化合物”[chalco-compound], “二酸化ケイ素”[silicon dioxide], “芳香族ポリアミン”[aromatic polyamine], “無機系化合物”[inorganic chemical compound]</p>
012	x	<p>“発光強度”[luminescence intensity], “発光効率”[luminescence efficiency], “歩留まり”[process yield], “白色発光”[white luminescence], “コスト”[cost], “光量”[light intencity], “光出力”[light output], “低電力”[low power], “発光パターン”[light pattern]</p>
012	y	<p>“窒化物半導体”[nitride semiconductor], “インジウム”[indium], “窒化ガリウム”[nitride gallium], “亜鉛酸化物”[zinc oxide], “ダブルヘテロ構造”[double-hetero structure]</p>
024	x	<p>“コスト”[cost], “効率”[efficiency], “耐久性”[durability], “安定” [stability]</p>
024	y	<p>“固体高分子”[solid giant molecule], “固体イオン導電体”[solid ionic conductor], “固体電解質”[solid electrolyte], “固体高分子電解質”[solid giant molecule electrolyte], “高分子固体電解質”[giant molecule electrolyte], “メタノール”[methanol], “親水性物質”[hydrophilia material], “黒鉛”[graphite], “フッ素”[fuluorine]</p>
025	x	<p>“曇”[fog], “曇り”[fog], “汚”[stain], “水滴”[water drop], “防曇”[defogging], “防露”[anti-moisture], “防汚”[anti-stain], “弾き現象”[flipping phenomenon], “硬度”[hardness], “透明”[transparence], “耐摩耗性”[anti-wearing-out]</p>
025	y	<p>“酸化チタン”[titanium oxide], “複合酸化物”[oxide compound], “金属酸化物”[metal oxide], “酸化珪素”[silicon oxide], “酸化モリブデン”[molybdic oxide], “酸化タングステン”[tungsten oxide], “撥水性フッ素樹脂”[water-repellent fluorocarbon resin], “無定型シリカ”[amorphous silica], “光触媒”[light catalyst]</p>

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

We would like to thank the evaluators who evaluated and gave useful comments for our results.

The NTCIR3 patent data collection was used in our research.

A part of this research is supported by a grant from the Artificial Intelligence Research Promotion Foundation.

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.