

特許請求項読解支援のための「発明の詳細な説明」との自動対応付け

新森 昭宏[†] 奥村 学^{††}

特許明細書には，特許請求項と「発明の詳細な説明」が記述される．特許請求項は特許明細書において最も重要な部分であるが，構成的または結合的に記述されるため，可読性が低い．「発明の詳細な説明」と，特許請求項を対応付けることにより，(1) 特許請求項に対する作用（機能）と効果を明確化する，(2) 特許請求項の重要箇所を明確化する，(3) 特許請求項で使われている表現に関する言い換えを取得する，等の効果が得られ，特許請求項の読解支援につながる．本稿では，特許請求項を構造解析し，その結果を用いて，「用言文節を起点としたローカルアラインメント」を行うことにより，「発明の詳細な説明」との対応付けを行う手法を提案する．NTCIR3の特許データコレクションからランダムに抽出した100件のうち88件の特許明細書を対象として評価を行い，その有効性を確認した．

キーワード: 特許請求項, 読解支援, 対応付け, ローカルアラインメント

Aligning Patent Claims with the “Detailed Description” for Readability

AKIHIRO SHINMORI[†] and MANABU OKUMURA^{††}

Patent specifications consist of patent claims and detailed descriptions. While patent claims are the most important part of patent specifications, they are compositionally or combinationally described and difficult to read. By aligning patent claims with the “detailed description”, (1) the functions and the effects of the claim can be clarified, (2) the important elements in the claims can be identified, or (3) paraphrases for the expressions in the claim can be obtained. In this paper, we propose a method to align patent claims with the “detailed description” by analyzing the structure of claims to get core elements of claims and by doing local alignments starting from word blocks including declinable words. By using 88 patent specifications out of 100 which were randomly picked up from the NTCIR3 patent data collection, the effectiveness of the method is demonstrated.

KeyWords: *patent claim, reading support, alignment, local alignment*

1 はじめに

特許行政の電子化がいち早く推進された結果，特許出願の内容を記述する「特許明細書」の電子データは既に膨大に蓄積されている．さらに，年間40万件以上と言われる新規出願に伴い，

[†] インテック・ウェブ・アンド・ゲノム・インフォマティクス, INTEC Web and Genome Informatics Corp.

^{††} 東京工業大学精密工学研究所, Precision and Intelligence Laboratory, Tokyo Institute of Technology

データ量は日々増加している．特許明細書は研究論文と同様の技術文書であり，これを有効活用するためのテキスト処理・自然言語処理技術に対する潜在的ニーズは大きい．

特許明細書は一種の法的文書であり，以下のような構造を持っている．

- 特許明細書
 - － 発明の名称
 - － 特許請求の範囲
 - － 発明の詳細な説明
 - * 発明の属する技術分野
 - * 従来技術
 - * 発明が解決しようとする課題
 - * 課題を解決するための手段
 - * 発明の実施の形態
 - * 発明の効果
 - － 図面の簡単な説明
- 図面
- 要約

最も重要な箇所である「特許請求の範囲」には，「特許請求項」（クレーム）が列挙される．特許請求項の記述は，1文で発明内容を記述するという制約と，独特の記述スタイルにより，専門家以外の人にとっては極めて読みにくいものになっている（新森，奥村，丸川，岩山 2004）．

特許法は，「特許請求の範囲」の記載について，「特許を受けようとする発明が発明の詳細な説明に記載したものであること」を規定している．「発明の詳細な説明」については，「その発明の属する技術分野における通常の知識を有する者がその実施をすることができる程度に明確かつ十分に記載しなければならない」としている．「発明の詳細な説明」は，通常の技術文書と同様の記述スタイルで記述されるため，比較的読みやすいが，記述量が多く，ポイントを絞りにくい．

以上を背景として，本稿では特許請求項の読解支援を目的とした，「発明の詳細な説明」との対応付け手法を提案する．2章では，このような対応付けを行うことの意義について説明する．次に，3章では，対応付けのための具体的な手法について説明する．4章では，NTCIR3の特許データコレクション（岩山，藤井，高野，神門 2001）からランダムに抽出した100件のうち88件の特許明細書を対象とした評価と，それに対する考察について報告する．5章では，関連研究について述べる．

2 特許請求項と「発明の詳細な説明」との対応付け

2.1 対応付けの意義

特許請求項と「発明の詳細な説明」との対応付けを行うことには、以下のような意義がある。

- (1) 特許請求項に対する作用（機能）と効果を明確化する
- (2) 特許請求項の重要箇所を明確化する
- (3) 特許請求項で使われている表現に関する言い換えを取得する

(1) に関しては、特許請求項の記述において「発明を明確に記述すること」という記述要件のもとで、「構成的」または「結合的」な表現が使われることが多く、「機能的」または「作用的」な表現はあまり使われないという事情に起因している（葛西泰二 1999；竹田和彦 2000；田辺徹 2000）。実際、日本の特許においては「から構成された」や「からなる」という表現が、アメリカの特許においては”comprising”や”consisting of”という表現が多用されている。構成的または結合的に記述される特許請求項は、通常の技術論文や技術解説書には見られないスタイルであり、当該技術分野の技術者が読解しようとした場合に問題となる。その特許請求項で記述された発明が結局のところ、どのような作用（機能）と効果をもたらすのかが、特許請求項を読んだだけではわからないことが多いためである。このため、「発明の詳細な説明」との対応付けにより、それが明確化されることは意義が大きい。

(2) に関しては、特許請求項が複数の構成要素から構成されているという事情に起因している。特許請求項が複数の要素や処理の説明を1文の中に詰め込んだ形で記述されているためであり、その観点から、特許請求項の記述スタイルを以下の3つに分類することができる（新森他 2004）。

順次列挙形式 「～し、～し、～した、～」のように、処理を順次的に記述する形式である。方法の発明で用いられることが多いが、物の発明において用いられることもある。

構成要素列挙形式 「～と、～と、～とからなる、～」のように、構成要素を列挙する形で記述する形式である。主に、物の発明で用いられる。

ジェプソン（Jepson）的形式 「～において、」や「～であって、」などの文字列を用いて記述を前半部と後半部に分割し、前半部と後半部にはそれぞれ「名詞まとめ」とそれに対する修飾部が前置されている形式である。前半部では、公知部分（既に知られている内容）または前提条件を述べ、後半部では新規部分（この発明の特徴となる部分）または本論部分を記述する。

上記の分類において、順次列挙形式や構成要素列挙形式の場合、3つ以上の要素を含むものが多い。ジェプソン的形式の場合は明らかに、後半部が前半部より相対的に重要であると考えられるが、その部分がさらに3つ以上の要素から構成されることも多い。複数の構成要素から構成されている特許請求項について、「発明の詳細な説明」との対応付けにより重要要素が明確化できれば、その特許請求項の重要箇所を明確化することができる。

インクを吐出するための1つのオリフィスに対して複数の電気熱変換素子を備えたインクジェットヘッドを用い、被記録媒体に記録を行うインクジェット記録装置において、インクジェットヘッドにおいて、駆動すべき電気熱変換素子を選択することによりそれぞれ異なる吐出量の複数の吐出モードを可能とするヘッド駆動手段と、該ヘッド駆動手段がより少ない吐出量の吐出モードで吐出を行わせるとき、当該吐出量より多い吐出量の吐出モードで、記録に関与しない予備吐出を行う予備吐出手段と、を備えたことを特徴とするインクジェット記録装置。

図1 特許請求項の例 (特開平 10-16222 より引用)

以上の説明から明らかなように、本発明によれば、小液滴の吐出モードのときは、より多い吐出量の吐出モードで予備吐出が行われるため、小液滴吐出モードでは除去し難く、また、発生し易い増粘インクや混色インクを容易に除去することができる。

図2 図1の特許請求項に対応付けられる「発明の効果」節中の文

(3) に関しては、法律文体的な文体、かつ1文の中に詰め込んだ形で記述されるということでもわかりにくく表現されているものに対して、それと同義のわかりやすい表現が得られれば、特許請求項の可読性向上に役立てることができる。

対応付けの効果を例示するために、図1に示す特許請求項に対応付けられた「発明の効果」節中の文を図2に示す。下線を付けた部分が、対応付けられた箇所である。なお、図1の特許請求項をわかりやすくするために、新森らによる構造解析手法(新森他 2004)で構造解析した結果を、図3に示す。

図1と図2の対応付けにより、図1の特許請求項の効果が、図2で特許請求項と対応付けられた箇所の後の「ため、」に後続する箇所(「小液滴吐出モードでは除去し難く、また、発生し易い増粘インクや混色インクを容易に除去することができる。」の箇所)として得られる。そして、図2で特許請求項と対応付けられた箇所は、図1の特許請求項の作用(機能)を表していると考えることができる。また、図1の特許請求項はジェブソンの形式であるが、図2との対応付けにより、後半部の2番目の要素が重要要素であることが明確化される。さらに、図1と図2との対応付けにより、以下の言い換えを得ることができる。

- 多い吐出量の吐出モードで、記録に関与しない予備吐出を行う
- 多い吐出量の吐出モードで予備吐出が行われる

2.2 「発明の詳細な説明」からの、対応付け対象文の抽出

1章で述べたように、「発明の詳細な説明」は通常、「発明の属する技術分野」から始まり、「発明の効果」までの、6つの節を持つ。このうち、「発明の属する技術分野」、「従来の技

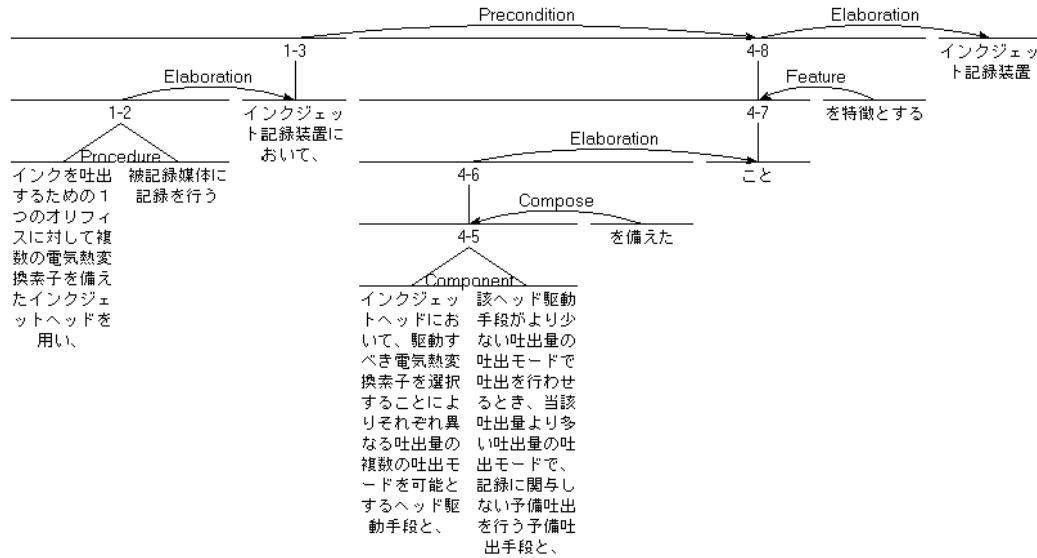


図 3 図 1の特許請求項を構造解析した結果 (RSTTool v2.7(OD'onnell 1997) を用いて表示)

術」, 「発明が解決しようとする課題」の3つの節は, 発明の背景情報や前提事項を説明する箇所であるため, それらの中に記述されている文は, 特許請求項との対応付けの対象とするには適さない. 残りの3つの節(「課題を解決するための手段」, 「発明の実施の形態」, 「発明の効果」)のうち, 我々は, 以下の理由により, まず「発明の効果」節に着目すべきであると考え.

特許出願者は通常, 「どんな発明を行ったか」に加えて, 「(発明によって)どんな作用(機能)がもたらされるか」と「(発明によって)どういう効果を生じることができるか」を特許明細書中に記述する. そして, 前節の例にみられるように, 「(発明によって)どんな作用がもたらされるか」に後続して, 「(発明によって)どういう効果を生じることができるか」を記述することが多い.

「発明の効果」節の記述特性を調査するために, NTCIR3の特許データコレクションの中から「レーザージェットプリンタに関する検索トピック」のデータである132件の特許明細書の「発明の効果」節から抽出した文の文末表現パターンと, 文中の読点直前の表現パターンを調査した. その結果を表1と表2に示す. なお, 表1と表2, 及び本稿の以降の記述では, パターンの表現に perlの正規表現を使用している. また, 表1では, それぞれのパターンについて, 「課題を解決するための手段」節と「発明の実施の形態」節中での出現割合を調査した結果も示している.

表1において, 上位5つはすべて, 効果を表す文の文末に出現する表現(以下では, 「効果表現」と称する)であると考えられ, その割合を合計すると, 69.8%になる. このことから, 「発明の効果」節には確かに, 効果を表現した文が多く記述されているということがわかる. また, 表2において, 第5位の「により」から第10位の「ため」までの表現は, 理由または原因と, 結果ま

表 1 「発明の効果」節から抽出した文の文末表現パターン

順位	パターン	「発明の効果」中での割合	「課題を解決するための手段」中での割合	「発明の実施の形態」中での割合
1	できる。	51.3 %	17.5%	9.9%
2	可能(に と)なる。	8.7 %	1.8%	0.9%
3	(得 え)られる。	4.8 %	1.2%	0.8%
4	可能である。	2.5 %	0.5%	1.2%
4	効果がある。	2.5 %	0.1%	0.0%

表 2 「発明の効果」節から抽出した文の読点直前の表現パターン

順位	パターン	割合
1	<動詞>	19.5 %
2	は	16.5 %
3	て	13.7 %
4	<名詞>	12.8 %
5	ば	5.6 %
5	により	5.6 %
7	ので	4.7 %
8	によれば	4.0 %
9	(従 したが)って	3.9 %
10	ため	3.4 %

たは効果をつなぐ表現であると考えられ、その割合を合計すると 21.6% になる。このことから、「発明の効果」節には、文中に「効果に対応する理由または原因」を表現した文が多く記述されているということが推察される。そして、「効果に対応する理由または原因」の部分が、この特許の作用（機能）であると考えられる。

外国での出願を翻訳した形で出願される特許などの場合は、「発明の効果」節を含まないものがある。また、「発明の効果」節の文として、文字通り、効果だけを述べた文しか含まないものもある。このような場合、「課題を解決するための手段」節に「（発明によって）どういう効果を生じることができるか」が記述されることがある。一方、「発明の実施の形態」節には、「実施例」という形で、発明内容が具体的かつ詳細に説明され、その中で効果が記述されることがある。しかし、「発明の実施の形態」節中で記述されるのは、実施時の細かな処置に対する効果であることが

多い。このため、それに前置する作用（機能）は、特許請求項に対応するものというよりも、より具体化された実施ステップに対応付けられるものが多い。

その他、特許明細書によっては、「作用」節が含まれるものがある。この節には文字通り、特許請求項に対する作用（機能）が記述されるが、そのような特許明細書はそれほど多くない。

上記を踏まえ、特許請求項との対応付けの対象とする文は、「発明の効果」節、（「作用」節が存在している場合は）「作用」節、「課題を解決するための手段」節、「発明の実施の形態」節の順に抽出するのが妥当であると考えられる。

3 対応付け手法

3.1 基本方針

対応付けにあたり、以下の基本方針を採用する。

- (1) 用言文節を起点としたローカルアラインメントをとる。

特許請求項の記述には正確性が求められるため、通常の記事では省略されるような指示詞を省略しないで記述すること、または「前記」や「当該」などの表現を用いることが多い。その他、1文で記述するという制約のために、通常の記事では別の文とするような説明記述を名詞の修飾語として用いることも多い。

対応付けは特許請求項に対応する作用（機能）を得るために行うこと、そして、作用（機能）に関する記述には通常、動詞や助動詞などの用言が用いられることを考えると、用言の部分に対応付けの起点とする必要がある。従って、用言の近くの局所的な類似性を探索するために、ローカルアラインメント (Gusfield 1997; 丸川, 岩山, 奥村, 新森 2002) を用いる。また、比較の基本単位としては、単語単位と文節単位のいずれかを選択可能であるが、我々は文節単位を用いることにした。

文節の比較においては、竹内ら (竹内 松本 2002) と類似した方法を用いることにした。すなわち、比較する文節をそれぞれ b_1 と b_2 とし、その類似度のスコアを $s(b_1, b_2)$ としたとき、以下のように定義する。

- b_1 と b_2 とが、（末尾に読点があった場合はそれを除去して）完全に一致した場合、
 $s(b_1, b_2) = 3$.
- b_1 と b_2 の主辞が用言で、その原型が同じ場合、または b_1 と b_2 が同一の名詞または記号に格助詞が後続する場合、 $s(b_1, b_2) = 2$.
- b_1 と b_2 が、共通の単語を含む場合、 $s(b_1, b_2) = 1$.

- (2) 特許請求項の構造解析結果を利用し、「対応付け対象要素」の集合を得る。

特許請求項の構造解析により、要素や処理を明確化することができる (新森他 2004) .

構造解析によってジェプソンの形式と判定された場合、後半部の要素だけを対応付けに用いる。順次列挙形式と構成要素列挙形式と判定された場合、各要素を対応付けに用いる。すなわち、特許請求項から「対応付け対象要素」の集合を得るために、構造解析の結果を利用する。

3.2 手順

- (1) 「発明の詳細な説明」章から、対応付け対象文を抽出する。その際、明確に他の特許請求項を対象として記述している文は除外する。
 - まず、「発明の効果」節から、表3に示す「効果表現」を含む文を検索し、その文が表4に示すような「直前の文に理由または原因があることを示唆する表現」から始まる場合は直前の文を、そうでない場合はその文を抽出する。なお、表3に示す「効果表現」は、表1中のものを基に手作業で表現を追加することで作成した。
 - 「発明の効果」節が存在しない場合、または「発明の効果」節から抽出した文と特許請求項の対応付けがうまくいかない場合、「課題を解決するための手段」節から、表3に示す「効果表現」を含む文を検索し、その文が表4に示すような「直前の文に理由または原因があることを示唆する表現」から始まる場合は直前の文を、そうでない場合はその文を抽出する。
 - 「課題を解決するための手段」節が存在しない場合、または「課題を解決するための手段」節から抽出した文と特許請求項の対応付けがうまくいかない場合、「発明の実施の形態」節から、表3に示す「効果表現」を含む文を検索し、その文が表4に示すような「直前の文に理由または原因があることを示唆する表現」から始まる場合は直前の文を、そうでない場合はその文を抽出する。
 - 明確に他の特許請求項を対象として記述されている文を除外するためには、たとえば第1請求項を対象として対応付けを行う場合、「請求項2」「請求項3」などの文字列を含む文を除外する。表4に示すような「直前の文に理由または原因があることを示唆する表現」から始まる場合は、直前の文が「請求項2」「請求項3」などの文字列を含む場合、その文を除外する。
- (2) 特許請求項と対応付け対象文との対応付けを行う。特許請求項を構造解析し、「対応付け対象要素」の列を得る。対応付け対象要素の列のうち、後ろの要素から順次的に、対応付け対象文との間で、「用言文節を起点としたローカルアラインメント」を実行し、各スコアを合計したものを全体のスコアとする。そして、スコアが最大となったものを選択する。対応付けアルゴリズムを、図4に示す。

表 3 対応付け対象文の抽出に使用する「効果表現」

(でき(る た) (に と)な(る った つた) (得 え)られ(る た) 可能である
効果 利点)がある あった あつた もたらされる もたらされた 生じる 生じた) 提供する
省け(る た) (図れ はかれ)(る た) (行 行な おこな)え(る た)
る(よう 様)にな(る った つた) され(る た)
((こと 事 おそれ 恐れ 必要)が し)(ない 無い なくなる 無くなる)(ものである)?(、 .)

表 4 直前の文に理由または原因があることを示唆する表現

((これ このこと この事)に(よ 因)(り れば って つて) (従 したが)って このため)(、 、)

入力：対応付け対象要素の列 (E), 対応付け対象文 (S)

出力：スコア (S_c), E 中で対応付けられた文字列の列 (Str_1), S 中で対応付けられた文字列の列 (Str_2)

1. $E =$ (対応付け対象要素)の列;
2. $S =$ 対応付け対象文;
3. $S_c = 0$;
4. while (E が空でない)
5. $e = E$ の最後の要素;
6. S を形態素解析し, 文節分割する. S の文節数を $|S|$, i 番目の文節を S_i とする ($i = 1, \dots, |S|$);
7. e を形態素解析し, 文節分割する. e の文節数を $|e|$, j 番目の文節を e_j とする ($j = 1, \dots, |e|$);
8. 2次元の表を作成し, ダイナミックプログラミングを行う;
9. $(|S| + 1) * (|e| + 1)$ の表 V を作成する;
10. S の文節列を行見出しに, e の文節列を列見出しにそれぞれ設定する;
11. $(\forall i, 0 \leq i \leq |S|) V(i, 0) = 0$;
12. $(\forall j, 0 \leq j \leq |e|) V(0, j) = 0$;
13. for $i = 1$ to $|S|$ do
14. for $j = 1$ to $|e|$ do
15. $V(i, j) = \max[V(i - 1, j) - 1, V(i - 1, j - 1) + s(S_i, e_j), V(i, j - 1) - 1]$;
16. end for
17. end for
18. for $j = |e|$ to 1 do
19. if (e_j が用言文節) then
20. last;
21. end if
22. end for
23. for $i = |S|$ to 1 do
24. if (S_i が用言文節) then
25. if ($V(i, j)$ の値が j 列の中で最大) then
26. $V(i, j)$ を起点とし, 次の e までの範囲でローカルアラインメントを探索する.
27. $S_c = S_c + V(i, j) -$ (ローカルアラインメント起点の V 値);
28. Str_1 の先頭に e で対応付けられた文字列を挿入;
29. Str_2 の先頭に S で対応付けられた文字列を挿入;
30. end if
31. end if
32. end for
33. $E = E \setminus \{e\}$
34. end while

図 4 対応付けアルゴリズム

表 5 特許請求項を構造解析して得られた要素（対応付け対象要素）と「発明の詳細な説明」から抽出した文（対応付け対象文）との間での対応付けの例

	...	当該吐出量より	多い	吐出量の	吐出モードで、	記録に	関与しない	予備吐出を	行う	予備吐出手段と、
以上の	0	0	0	0	0	0	0	0	0	0
説明から	0	0	0	0	0	0	0	0	0	0
明らかのように、	0	0	0	0	0	0	0	0	0	0
本発明に	0	0	0	0	0	0	0	0	0	0
よれば	0	0	0	0	0	0	0	0	0	0
小液滴の	0	0	0	0	0	0	0	0	0	0
吐出モードの	0	0	0	1	2	1	0	1	0	0
ときは、	0	0	0	0	1	0	0	0	0	0
より	0	0	0	0	0	0	0	0	0	0
多い	0	0	3	2	1	0	0	0	0	0
<u>吐出量の</u>	0	1	2	6	5	4	3	2	1	1
<u>吐出モードで</u>	4	3	2	5	9	8	7	6	5	4
<u>予備吐出が</u>	6	5	4	4	8	7	6	9	8	7
<u>行われる</u>	9	8	7	6	7	6	5	8	11	10
<u>ため、</u>	8	7	6	5	6	5	4	7	10	9
<u>小液滴吐出モードでは</u>	7	9	8	7	6	5	4	6	9	11
<u>除去し難く、また、</u>	6	8	7	6	5	4	3	5	8	10
<u>発生し</u>	5	7	6	5	4	3	2	4	7	9
<u>易い</u>	4	6	5	4	3	2	1	3	6	8
<u>増粘インクや</u>	3	5	4	3	2	1	0	2	5	7
<u>混色インクを</u>	2	4	3	2	1	0	0	1	4	6
<u>容易に</u>	1	3	2	1	0	0	0	0	3	5
<u>除去する</u>	0	2	1	0	0	0	0	0	3	4
<u>ことができる。</u>	0	1	0	0	0	0	0	0	1	3
	0	0	0	0	0	0	0	0	0	2

- 対応付けの例として、図 1の特許請求項を構造解析して得られた要素の一つを「対応付け対象要素」 e とし、図 2の文を「対応付け対象文」 S として、「用言文節を起点としたローカルアラインメント」を行った結果を、表 5に示す。表中で、下線部分が対応付けできた部分である。

4 評価と考察

4.1 評価

新森らによる特許請求項構造解析の論文（新森他 2004）において直接評価で用られた 100 件の特許明細書（NTCIR3 の特許データコレクションのうち 1999 年の公開特許公報から、公開日が 1999 年 1 月から 3 月までの 59,956 件の中からランダムに選択した 100 件）を用いて、本手法を評価した。第 1 請求項を対象として構造解析を行ったため、本評価でもそれぞれの特許明細書の第 1 請求項を対象として対応付けを行った。

発明の分野によっては、特許請求項と対応付けることのできる文を「発明の詳細な説明」に含まないものがあるということに留意する必要がある。たとえば、化学物質に関する発明の場合、化

学式だけが特許請求項に記載されることが多い。このような点を考慮し、「特許公報及び公開特許公報発行区分表」(特許庁編 2000)において、「III 化学・冶金・繊維」の部門に属するもの、つまり以下のIPC(International Patent Classification: 国際特許分類)のコードのものについては、評価の対象外とした。なお、特許庁が発表した1999年の年間出願データによれば、「III 化学・冶金・繊維」の部門の構成割合は、9.7%である。

無機化学 C01, C03-C06, C30

有機化学 A01N, A61K, C07

高分子化学 C08-C11, C14

冶金 B22F, C21-C25

繊維 A41-42, D

評価対象の100件の特許明細書のうち、上記の分類に属するものは、11件であった。残りの89件の特許明細書の中には、論文(新森他 2004)の時点でのプログラムを用いた第1請求項の構造解析が失敗したものを10件含んでいた。このうち、9件については、構造解析の間違ひは本手法の評価を行う上で本質的ではないもの(ジェブソンの形式の前半部での解析間違ひ等)、または構造解析の間違ひを人手で容易に修正可能なものであったため、本評価の対象として用いることにした。すなわち、1件を除いた88件について、評価対象とした。

以下の評価項目について、人手による評価を行った。なお、以下の評価項目の評価値算出においては、評価対象の特許明細書数を r 、特許請求項と対応付け可能な文として抽出されるべき文の数を t 、実際に抽出された文の数を e 、対応付けられた文の数を a 、対応付け結果が正しいものの数を c 、対応付けにおいて特許請求項の構成要素の一部だけが対応付けされたものの数を i 、対応付けにより言い換え表現が取得できたものの数を p 、対応付け対象文で対応付けられた文字列の末尾またはその直後に後続する形で理由または原因を表す表現(因果表現)が存在しているか、またはその文字列が文末にあるものの数を q としている。

対応付け対象文の存在率 特許請求項と対応付け可能な文が存在するかどうかを、 t/r により評価する。今回の場合は、 $r = 88$ である。

対応付け対象文の抽出率 特許請求項と対応付け可能な文が抽出できたかどうかを、 e/t により評価する。

対応付けの成功率 抽出した対応付け対象文に対して、本手法で対応付けができたかどうかを a/e により評価する。

対応付けの正解率 対応付け結果は、両文の意味を考慮して判定した場合、正しいかどうかを c/a により評価する。

構成要素の一部だけが対応付けされたものの割合 特許請求項の構成要素の一部だけが対応付けされたものの割合を i/c により評価する。この場合、対応付けされた構成要素が、特許請求項の重要箇所とみなすことができる。なお、ジェブソンの形式の場合に、後半部分のすべて

表 6 「対応付け対象文の存在率」と「対応付け対象文の抽出率」についての評価結果

項目	評価値
対応付け対象文の存在率	82/88 ≈ 0.93
対応付け対象文の抽出率	76/82 ≈ 0.93

の要素が対応付けされた場合は、前半部分の構成要素も含めた対応付けをもう一度実施し、後半部分の構成要素だけが対応付けに使われていることが確認できた場合、構成要素の一部だけが対応付けされたとみなす。

言い換え表現が取得できたものの割合 対応付けにより、特許請求項中の記述に対する言い換え表現が取得できたかどうかを p/c により評価する。全く同一の文字列の対が得られた場合は、言い換え表現の取得とはみなさない。また、一方が他方の部分文字列となっている場合も、言い換え表現の取得とはみなさない。

因果表現接続率 対応付け対象文で対応付けられた文字列の末尾またはその直後に理由または原因を表す表現（因果表現）が存在しているか、またはその文字列が文末にあるかどうかを q/c により評価する。

対応付けられた文字列の平均長 特許請求項と対応付け対象文のそれぞれについて、対応付けされた文字列の平均長

対応付けのベースラインとしては、以下の2つを設定した。

ベースライン 1 特許請求項の構造解析結果を全く使用せずに、特許請求項に出現する全ての用言文節を起点としたローカルアラインメントを（後ろから）順次的にとる。ただし、特許請求項に頻出する“を特徴とする”という表現の“する”は除外する。この場合は明らかに、上記の評価項目のうち、「構成要素の一部だけが対応付けされたものの割合」は評価対象外となる。

ベースライン 2 特許請求項の構造解析結果は使用するが、起点を用言文節にしないでローカルアラインメントを順次的にとる。

「対応付け対象文の存在率」と「対応付け対象文の抽出率」についての評価結果を表 6 に示す。それ以外の評価結果を表 7 に示す。

4.2 考察

前節で評価の対象外とした通り、本稿で提案している手法は、「III 化学・冶金・繊維」の分野に属する特許明細書の特許請求項の読解支援に貢献することはできないと考えられる。しかし、それ以外の分野の構成割合が約 90%であることを考えると、本稿で提案している手法には十分な実用性があると言える。

表 7 その他の評価項目についての評価結果

項目	本手法	ベースライン 1	ベースライン 2
対応付けの成功率	75/76 \approx 0.99	74/76 \approx 0.97	76/76 = 1.00
対応付けの正解率	66/75 \approx 0.88	55/74 \approx 0.74	56/76 \approx 0.74
構成要素の一部だけが対応付けされたものの割合	47/66 \approx 0.71	(評価対象外)	(評価対象外)
言い換え表現が取得できたものの割合	45/66 \approx 0.68	32/55 \approx 0.58	(評価対象外)
因果表現接続率	33/66 \approx 0.50	24/55 \approx 0.44	24/56 \approx 0.43
対応付けられた文字列の平均長 (特許請求項)	35.9 文字	10.1 文字	37.7 文字
対応付けられた文字列の平均長 (対応付け対象文)	34.1 文字	9.9 文字	34.3 文字

表 6 によれば、「対応付け対象文の存在率」は約 0.93 であり、「III 化学・冶金・繊維」の部門を除いた特許明細書の約 93% に、特許請求項の対応付け対象となる文が存在すると言える。対応付け対象文が存在しなかった特許明細書は 6 件であった。そのうち 3 件は個人発明家による特許出願 (= 出願人が個人名であるもの) であり、通常の特許明細書とは少し異なるスタイルで記述されているものであった。また、2 件は、特許請求項の作用 (機能) を記述することなく、「発明の効果」を記述しているものであった。残り 1 件は、「III 化学・冶金・繊維」の部門ではないが、特許請求項に化学式を記述しているものであった。

「対応付け対象文の抽出率」は約 0.93 であり、本稿で提案している手法により、対応付け対象文の約 93% が抽出できていると言える。抽出された対応付け対象文 76 件のうち、63 件 (約 83%) は「発明の効果」節から、8 件 (約 10%) は「課題を解決するための手段」節から、5 件 (約 7%) は「発明の実施の形態」節から抽出された。対応付け対象文を抽出できなかったのは 6 件であった。そのうち、3 件は「直前の文に理由または原因があることを示唆する表現」の漏れであった。具体的には、「このように」という表現から文が始まっていた場合が 2 件、「この発明の半導体装置によれば」という表現から文が始まっていた場合が 1 件であった。残り 3 件は、「効果表現」の漏れであった。具体的には、「効果を奏する」という表現が 2 件、「ようにしている」という表現が 1 件であった。これらの表現を対応付け対象文の抽出プログラムに追加することで、「対応付け対象文の抽出率」を向上させることができる。

「言い換え表現が取得できたものの割合」は、45 件 (約 68%) であった。本手法により得られた言い換え表現の例を付録に示す。言い換えのパターンを分類すると、以下の通りとなる。

- 「上記」「前記」、図中記号などを除去したもの
- 動詞の能動態を受動態にしたもの、またはその逆
- 動詞を「テイル形」または「テアル形」にしたもの
- 動詞に「ようにした」または「ようにしている」を後続させたもの

表 8 対応付け失敗の分析

番号	説明	件数
1	特許請求項の作用（機能）を全く別の表現で表現し、その表現中の文字列が偶然に構成要素中の文字列と対応付けられてしまった場合	5
2	特許請求項の構成要素がさらに下部構造をもっており、本手法が使用している構造解析手法で検出した構成要素中で最も後方に出現する用言文節は、その構成要素の作用（機能）を表現するものとしては不適切である場合	1
3	他の請求項に対して対応付けられるべき文と対応付けてしまった場合	2
4	特許請求項中の表現順と対応付け対象文での表現順とが入れ替えになっているため対応付けが部分的になってしまった場合	1

- 上記以外の言い換え

対応付けに失敗したものの9件について、その分類分けとそれぞれの件数を、表8に示す。表8のうち、2番は特許請求項の構造解析手法の精緻化、3番は対応付け対象文の抽出処理の改善、4番は対応付け手法の改善で、対応可能であると考えられる。

「因果表現接続率」は、本手法の場合約0.50であり、それほど高くないが、その全ての場合において、因果表現の後に、効果を表す表現が続いていた。つまり、これらの場合は、特許請求項に対する作用（機能）と効果を明確化できたことになる。該当しなかったものとしては、言い回しを変えた形で表現されている場合、補足的な表現や付属的な表現が後続する場合などがあつた。

本手法とベースライン1を比較すると、「対応付けの成功率」、「対応付けの正解率」、「言い換え表現が取得できたものの割合」、「因果表現接続率」の全てにおいて、本手法がベースライン1を上回った。そして、「対応付けられた文字列の平均長」については、特許請求項側と対応付け対象文側の両方において、本手法がベースライン1を大幅に上回った。ベースライン1は、全ての用言文節を起点とするため、間違つた対応付けをする率が高くなっている。また、対応付けられる文字列が短いものになりがちである。

本手法とベースライン2を比較すると、「対応付けの成功率」でベースライン2が少し上回つたが、「対応付けの正解率」と「因果表現接続率」ではベースライン2が少し下回る結果となつた。また、「対応付けられた文字列の平均長」は、特許請求項側と対応付け対象文側の両方において、ベースライン2の方が本手法よりも少し大きな値となつた。ベースライン2は本手法に比べると、不適切な対応付けを行つてしまう傾向があり、かつ少し長い範囲の対応付けを行つてしまうことが多いと言える。たとえば、表5の例の場合は、以下の下線部のような対応付け結果を得てしまうことになる。

対応付け対象要素 e 該ヘッド駆動手段がより少ない吐出量の吐出モードで吐出を行わせるとき、

当該吐出量より 多い吐出量の吐出モードで、記録に関与しない予備吐出を行う予備吐出手段と、

対応付け対象文 S 以上の説明から明らかなように、本発明によれば、小液滴の吐出モードのときは、より 多い吐出量の吐出モードで予備吐出が行われるため、小液滴吐出モードでは除去し難く、また、発生し易い増粘インクや混色インクを容易に除去することができる。

5 関連研究

近年、要約や言い換えに関する研究を目的として、単一言語のコーパスについての対応付けの研究が盛んに行われている (Marcu 1999; Jing 2002; Barzilay and Elhadad 2003) .

特許明細書を単一言語コーパスとして対応付けの研究を行ったものには、(Murata and Isahara 2002; 丸川他 2002) がある。丸川ら (丸川他 2002) は、特許請求項と「発明の詳細な説明」との対応付けにおいて、交差の問題を、複数のローカルアラインメントを同定することで解決する手法を提案している。

村田ら (Murata and Isahara 2002) は、特許請求項と「発明の詳細な説明」との対応付けに、diff コマンドを用いた手法を提案している。しかし、この手法は、複数の特許請求項を持つ特許明細書について、実施例が各特許請求項の対応付け箇所から構成されていること、その箇所は特許請求項の出現順に存在していることなど、を前提している。この前提を満たす特許明細書は、現実にはそれほど多くない。また、特許請求項が1つしか存在せず、それに対して膨大な実施例が記述されている特許明細書の可読性向上については何の貢献もしない。

Sheremetyeva により、英語の特許請求項の解析に関する研究が報告されている (Sheremetyeva 2003) . 特許請求項の解析結果を用いて特許請求項の可読性を向上させる可能性も示唆されているが、「発明の詳細な説明」を利用するということには及んでいない。

6 まとめ、及び今後の課題

特許請求項と「発明の詳細な説明」との対応付けを行うことの意義を述べ、特許請求項の構造解析結果を用いた対応付け手法を提案した。NTCIR3の特許データコレクションからランダムに抽出した100件のうち88件の特許明細書を対象とした評価により、その有効性を確認した。

本稿で提案している対応付け手法は、新森らが提案した特許請求項構造解析 (新森他 2004) の結果を利用するため、「独立形式請求項」(他の特許請求項を引用せずに独立して記述される形式の請求項) だけを対象とすることになる。特許明細書には、複数の特許請求項が記述され、またそのうちいくつかは他の特許請求項を引用する形で記述される「引用形式請求項」となることが多い。引用形式請求項については、引用元の特許請求項の記述内容を参照した処理が必要となる。

更に今後は、特許明細書の読解支援という立場で研究を進める必要があると考えている。特許

明細書読解支援システムを実現するには、以下のような研究課題に取り組む必要がある。

複数の特許請求項の関係解析，及び「発明の詳細な説明」との対応付け 特許明細書には複数の特許請求項が記述されていることが多く，その相互関係がわかりにくいものが多いため，その関係解析を行う。また，「発明の詳細な説明」との対応付けを行う。

特許明細書の特性・構造に適した自動要約手法 特許明細書には，特許明細書作成者によって作成された「要約」が添付されている。しかし，法的に最も重要な箇所である特許請求項の記述内容を含んでいないことがあるなど，その内容は必ずしも信頼できるものではない。このため，特許明細書の特性・構造に適した形で自動要約を行い，特許明細書の読解支援に役立てる。特許検索にも有効利用できる可能性が大きい。

「特許マップ」の作成 関連する複数の特許明細書について，その相互関係の解析を行い，2次元の表などの形で視覚化する (Fujii, Iwayama, and Kando 2004)。

謝辞

本研究では，NTCIR3 ワークショップで配布された「特許データコレクション」を使用しています。また，奈良先端科学技術大学院大学松本研究室で開発・公開された「茶釜」，及び，工藤拓博士が開発・公開された「南瓜」を使用しています。さらに，Michael OD'onnell 博士が開発・公開された RSTTool を使用しています。本研究の一部は，財団法人 人工知能研究振興財団による「人工知能研究助成」の支援を受けています。

参考文献

- Barzilay, R. and Elhadad, N. (2003). "Sentence Alignment for Monolingual Comparable Corpora." In *Proceedings of the Eighth Conference on Empirical Methods in Natural Language Processing*.
- Fujii, A., Iwayama, M., and Kando, N. (2004). "Overview of Patent Retrieval Task at NTCIR-4." In *Proceedings of the Fourth NTCIR Workshop on research in information Retrieval, Automatic Text Summarization and Question Answering*. NII (National Institute of Informatics).
- Gusfield, D. (1997). *Algorithms on Strings, Trees and Sequences*. Cambridge University Press.
- Jing, H. (2002). "Using Hidden Markov Modelling to Decompose Human-Written Summaries." *Computational Linguistics*, **28** (4), 527-543.
- Marcu, D. (1999). "The automatic construction of large-scale corpora for summarization research." In *Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 137-144.
- Murata, M. and Isahara, H. (2002). "Using the Diff Command in Patent Documents." In *Proceedings of the Third NTCIR Workshop on research in information Retrieval, Automatic*

- Text Summarization and Question Answering*. NII (National Institute of Informatics).
- OD'onnell, M. (1997). "RST-Tool: An RST Analysis Tool." In *The 6th European Workshop on Natural Language Generation*.
- Sheremetyeva, S. (2003). "Natural Language Analysis of Patent Claims." In *Proceedings of the Workshop on Patent Corpus Processing (ACL 2003 Post-Conference Workshop)*, pp. 66-73.
- 葛西泰二 (1999). 特許明細書のクレーム作成マニュアル. 工業調査会.
- 新森昭宏, 奥村学, 丸川雄三, 岩山真 (2004). "手がかり句を用いた特許請求項の構造解析." 情報処理学会論文誌, 45 (3), 891-905.
- 岩山真, 藤井敦, 高野明彦, 神門典子 (2001). "特許コーパスを用いた検索タスクの提案." 情報処理学会研究会報告 - 情報学基礎, 63-7号, pp. 49-56.
- 竹田和彦 (2000). 特許の知識【第6版】. ダイヤモンド社.
- 田辺徹 (2000). 英文特許の常識. 工業調査会.
- 特許庁編 (2000). 特許出願の手引き (第29版). 発明協会.
- 丸川雄三, 岩山真, 奥村学, 新森昭宏 (2002). "ローカルアラインメントを用いたテキスト間の柔軟な対応 付け." 情報処理学会研究会報告 - 自然言語処理, 151-4号, pp. 23-28.
- 竹内和広 松本裕治 (2002). "自動文節対応付けを用いた要約中の文再構成操作の調査." 自然言語処理, 9 (3), 87-108.

付録

得られた言い換え表現の例を以下に示す .

特開平 11-3065 より

- 周期に応じて前記遅延制御手段の1ステップ当たりの遅延時間を可変する
- 周期に応じて可変するようにした

特開平 11-4202 より

- 発生する自身のステータスデータを、前記処理データと共に直接多重化する
- 発生する自身のステータスデータを、複数の機器から入力される処理データと共に当該多重化装置により直接多重化されるので、

特開平 11-4737 より

- 突出片が前記フレームの凹溝に着脱自在に嵌合された位置で前記フレームの端末同士が接合状態となるようにした
- 突出片がフレームの凹溝に嵌合した位置でフレームの端末同士が接合状態となる。

特開平 11-19742 より

- 編成機本体の回転ドラムと結合して同方向へ一体に回転させ、

- 編成機本体の回転ドラムと一体に回転させるようにしたので、

特開平 11-22818 より

- 所定の回転部材の回転数が予め定めた回転数以上に増大しない油圧に設定する
- 所定の回転部材の回転数が予め定めた値以上に増大しないように

特開平 11-25597 より

- 記録媒体へ記録された後の再生の制限を示す制限情報とが多重化され送信される
- 記録媒体への記録再生の制限を示す制限情報とが多重で送信される

特開平 11-28203 より

- X線検出器の内の一部の列の検出器を、散乱線検出器として用いる
- X線検出器列の一部のX線検出器列を散乱X線検出器として用いる

特開平 11-52471 より

- 分類の内容を表示し、選択された分類の中に含まれる階層構造でなる内容をスクロール表示するように制御する
- 分類の中に含まれる階層構造でなる内容をスクロール表示させるようにしているの
で、

略歴

新森 昭宏: 1983年京都大学理学部卒業. 1990年コロラド大学コンピュータサイエンス学科修士課程修了. 1983年(株)インテック入社. 2000年, インテック・ウェブ・アンド・ゲノム・インフォマティクス(株)に移籍. 同社アドバンスト・リサーチ・グループ ジェネラルマネージャ. 技術士(情報工学部門). 情報処理学会, 言語処理学会, 人工知能学会, ACL, ACM, AAAI 各会員.

奥村 学: 1962年生. 1984年東京工業大学工学部情報工学科卒業. 1989年同大学院博士課程修了. 同年, 東京工業大学工学部情報工学科助手. 1992年北陸先端科学技術大学院大学情報科学研究科助教授, 2000年東京工業大学精密工学研究所助教授, 現在に至る. 工学博士. 自然言語処理, 知的情報提示技術, 語学学習支援, テキストマイニングに関する研究に従事. 情報処理学会, 人工知能学会, AAAI, 言語処理学会, ACL, 認知科学会, 計量国語学会各会員.
oku@pi.titech.ac.jp, <http://oku-gw.pi.titech.ac.jp/~oku/>.

(年 月 日 受付)

(年 月 日 再受付)

(年 月 日 採録)