

特徴語句の修飾関係を考慮した 特許請求の範囲の構造解析に関する研究

檜山武浩† 田中成典‡ 杉之原亮† 持永大輔† 増満光‡

関西大学大学院総合情報学研究科† 関西大学総合情報学部‡

1. はじめに

近年、我が国では、特許をはじめとする知的財産に関する法整備を進めている。特許法についても毎年時代に沿った改正がされている。しかし、特許業務の基本作業である特許調査については、その特性上簡略化されず多くの費用が費やされているという問題がある。これは、発明の内容を定義する特許請求の範囲の文書書式が複雑であることと、特許出願件数が年間 40 万件を超えることに起因する。この問題に対して、既存研究では、特許文書の主題分析による文書分類手法[1]、検索手法[2]やそれに準じる解析手法[3][4]などが提案され、関連特許の調査対象を絞り込むことには成功している。しかし、特許文書に関する既存研究は、可読性向上を目的としていないため、特許文書の内容把握には、特許調査者の読解力に依存しているという問題がある。そこで本手法では、特許請求の範囲の特徴[5]を使用することで、特徴語句に係わる修飾関係を解析し、特許請求の範囲の修飾関係を明確化するための解析手法を提案する。

2. 研究の概要

本手法では、公開特許文書を入力データとして、特許請求の範囲の特徴に注目した解析を行い、名詞句の関係図生成を試行する手法を提案する。特許請求の範囲の特徴とは、特許請求の範囲が複数の項で発明を示す名詞句を説明していること、各請求項は 1 つの名詞句を修飾している名詞句であることと各請求項の形式に特定の特徴語句が修飾に使われていることである。本システムの概要を図 1 に示す。本システムは、1) 請求項の抽出機能、2) 特許請求の範囲の解析機能、3) 特許請求の範囲の可視化機能により構成される。

Reserch on Structure Analysis of Japanese Patent Document with Dependency Relation of Characteristic Words

†Takehiro Kashiyama, Ryo Suginoara, Daisuke Mochinaga
Graduate School of Informatics, Kansai University, 2-1-1
Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

‡Shigenori Tanaka, Hikaru Masumitsu
Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-
cho Takatsuki-shi. Osaka 569-1095. Japan

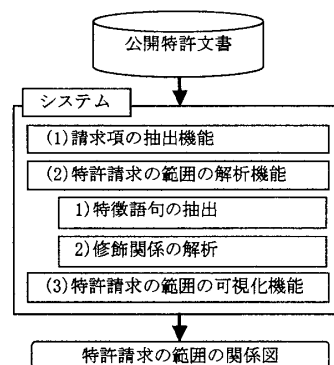


図 1 システムの概要

2. 1 請求項の抽出機能

本機能では、公開特許文書から特許の実態を定義する特許請求の範囲を抽出し、さらに請求項ごとにデータを分割する。

2. 2 特許請求の範囲の解析機能

本機能では、特許請求の範囲に対して、修飾関係判定に必要な情報を付加するため、特徴語句の抽出と修飾関係の解析を行う。

2. 2. 1 特徴語句の抽出

特徴語句の抽出では、正規表現を用いて特徴語句を抽出する。特徴語句とは、「前記」や「請求項～に記載」の既出の名詞句を参照する語句と「ジェプソン形式」の「～において～を特徴とする」や「～に当たり～を特徴とした」などの請求項の形式を特徴づける語句を指す。

2. 2. 2 修飾関係の解析

修飾関係の解析では、抽出した特徴語句に係わる修飾関係を解析する。解析は、特徴語句そのもの、特徴語句前後の品詞や文書内に記述されている名詞句などを対象とする。解析対象から文書中での役割を判定し、それぞれの文字列に対して解析結果を属性として付加する。

2. 3 特許請求の範囲の可視化機能

本機能では、特許請求の範囲の解析機能で解析した修飾関係について、請求項同士の参照関係や請求項内部の参照関係を複数のレイヤを用いて可視化し、可読性向上のための可能性を探る。

3. システムの実証実験と考察

実証実験では、公開特許から本手法を用いて特許請求の範囲の関係図を生成し、特許請求の範囲に記されている内容と同一であることを目視で比較し調査する。確認項目は、「前記」については参照先の名詞句、「請求項～に記載」については参照先の請求項番号と参照先の名詞句、「構成要素列挙形式」と「ジェプソン形式」については修飾先の名詞句とする。なお、化学分野の特許で使用される「マーカッシュ形式」には、今回対応していない。また、今回の入力データには、独立行政法人工業所有権情報・研修館が運営している特許電子図書館の公開特許データを使用する。

3. 1 実証実験

実証実験として、公開特許から無作為に選択した 25 件の特許データを用いて修飾関係を解析し、解析結果をもとに關係図を生成する。このとき、本手法の解析結果が、特許請求の範囲で定義されている修飾関係と同一であることを確認し、確認項目ごとの出現回数、正答率を調査した。さらに解析結果をもとに請求項と名詞句に対して参照関係を可視化する。このとき、「すべて」に対して参照するときは三角矢印、「いずれか」に対しては菱形矢印を使用する。

3. 2 結果と考察

まず、実証実験の結果として、各項目の出現回数と正解率を表 1 に示す。

表 1 解析精度の評価

確認項目	出現回数	正解率
参照	前記	739 91.74%
	請求項～に記載	147 93.19%
形式	構成要素列挙形式	63 90.47%
	ジェプソン形式	99 98.99%

表 1 の結果により、本手法が特許請求の範囲をほぼ再現していることが確認できた。正解できなかった主な要因としては、今回作成した正規表現や「IF-THEN」ルールの不足であった。他に、助詞の連体化「の」が名詞句内に存在するときに名詞句の切れ目が曖昧となってしまう、請求項に記された内容と違った名詞句を構成している場合があった。これは、構成要素列挙形式の「備えた」や「構成された」などが修飾する名詞句で多く確認された。この点については、現在の「IF-THEN」ルールを改める必要があると考える。次に、例として「分散発電システムとそれを利用可能なエネルギー管理システム：特許公開 2004-6404」を対象としたときの請求項の關係図を図 2 に示す。

分散発電システムとそれを利用可能なエネルギー

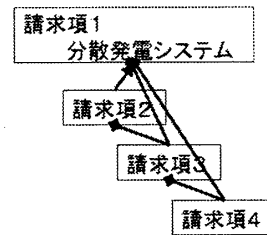


図 2 特許請求の範囲の關係図

図 2 の關係図は、請求項 1 を請求項 2 以降の請求項が参照して「分散発電システム」を説明していることを示している。このように文書全体の関係を明確化することで、理解度は向上すると考えられる。また、これは名詞句同士の關係についても同様の効果が期待できると考える。しかし、今回は、可読性向上を考慮した出力を行っていないため、現状では可読性向上には至っていないと考える。可視化のときのユーザビリティについて考慮する必要がある。

4. おわりに

本研究では、特許請求の範囲の解析と關係図の生成を実現した。実証実験では、25 件の特許文書を用いて修飾関係の解析精度を確認し、可視化による可読性向上の可能性を示した。今後の課題としては、可読性向上のための可視化手法とそれに関する解析手法提案を目指す。

参考文献

- [1] 間瀬久雄, 辻洋, 絹川博之, 石原正博: 特許テーマ分類方式の提案とその評価実験, 情報処理学会論文誌, 情報処理学会, Vol. 39, No. 7, pp. 2207-2216, 1998. 7.
- [2] 高木徹, 藤井敦, 石川徹也: 検索質問の主題分析に基づく類似文書検索と特許検索への応用, 情報処理学会論文誌, 情報処理学会, Vol. 46, No. 4, pp. 1074-1081, 2005. 4.
- [3] Marcu, D.: The Rhetorical Raising of Unrestricted Texts: A Surface-based Approach, Computational Linguistics, Vol. 26, No. 3, pp. 395-448, 2000. 5.
- [4] Sheremetyeva, S., Nirenburg, S.: Knowledge Elicitation for Authoring Patent Claims, IEEE Computer, Vol. 29, Issue 7, pp. 57-63, 1996. 7.
- [5] 新森昭宏, 奥村学, 丸川雄三, 岩山真: 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, 情報処理学会, Vol. 45, No. 3, pp. 891-905, 2004. 3.