

## 特許書類の可視化とハイパーテキスト化 Visualization and Transformation to Hypertext of Patent documents

高木 慎也<sup>†</sup> 新森 昭宏<sup>†</sup>

Shinya Takagi Akihiro Shinmori

### 1. はじめに

我が国においては近年、国家レベルの知財戦略が打ち出され、広い範囲の企業関係者が特許に関わらざるを得ない状況が生まれている。

それ故に知財専門家だけでなく、研究開発に携わる者でも、関連研究の調査等の理由により、特許書類を読むことが求められているのが現状である。

特許書類の中でも「特許請求の範囲」は一文が長かったり、独特な言い回しが多かったりと、特に読解が困難であることが知られている。

そこで我々は、主に技術者に対して、特許書類の読解を支援するシステムの開発を行っている。以下上記システムを、特許書類読解支援システムと呼称する。本稿では、特許書類読解支援システムの概要と機能の詳細について説明する。

### 2. 特許書類読解支援システム

#### 2.1 概要

特許書類読解支援システムは、「特許請求の範囲」の見える化を実現するシステムであり、具体的には、「注釈付きクレームツリー表示機能」と「特許請求項の細粒度解析機能」を持つシステムである。

まず特許書類テキストファイルを入力し、内部に組み込んである「特許書類解析エンジン」が解析を行い、解析結果をXMLファイルとして保存する。このXMLファイルのフォーマットとしては、PML(Patent Markup Language)<sup>[1]</sup>を使用している。そして、保存したXMLファイルを入力し、別のツールが「注釈付きクレームツリー」と「特許請求項の細粒度解析」を出力する。画面のイメージは図1の通りである。

図1：特許書類読解支援システムの出力画面のイメージ（特開 2004-101107 を表示）

図1に示す通り、解析結果を表示する際は、ブラウザで複数のフレームを用意し、複数の画面を表示するようにしている。具体的には、特許書類本体、請求項切り替えボタン、「注釈付きクレームツリー」と「特許請求項の細粒度解析」の表示である。その中で「注釈付きクレームツリー」と「特許請求項の細粒度解析」の表示は、フレーム内においてFlashで出力するようにしている。Flashを用いたのは、イメージの表現力に於いて優れていたからである。また、Flash内部のタブによって「注釈付きクレームツリー」と「特許請求項の細粒度解析」のそれぞれの結果を切り替えるようにしている。

## 2.2 注釈付きクレームツリー

特許請求項には、他の特許請求項とは独立に記述されるものと、他の特許請求項を引用して記述されるもののが存在するが、その引用関係は複雑となってしまうことが多い。「注釈付きクレームツリー」とは、まず特許請求項間の関係（引用形式特許請求項、外的付加、内的付加など）を、特許請求の範囲の記述から言語解析し<sup>[2][3][4]</sup>、更に各特許請求項で定義されている構成要素名を取得し、特許請求項間の多様な関係や請求項内の構成要素を、一目見て分かるように木構造で表したものである(図2)。引用形式請求項は、それ以前に記述した請求項を引用することになっているので<sup>[5]</sup>、請求項同士が枝で結ばれているものは、後の請求項が前の請求項を引用していることを表している。

「注釈付きクレームツリー」では、請求項の背景色を変えることで、灰色が独立形式請求項、青色が引用形式請求項(内的付加)、赤色が引用形式請求項(外的付加)、桃色が引用形式請求項(その他)として見分けることができるようにしている。

また、請求項のボックスの中には構成要素または内的付加の場合の制限対象が表示されており、白丸のものがジェブソン型のものにおける前提部に存在するもの、黒丸のものが本論部に存在するものを表している。

結果を出力する際は、上記の構成要素と制限対象を表示したツリーと、表示しないツリーの二種類の結果を出力する。図2と図3は要素を表示したツリーである。図2の場合、請求項1は独立形式請求項であり、請求項2は内的付加の引用形式請求項で請求項1を引用しており、請求項3は内的付加の引用形式請求項で、請求項1と2を引用していることを表している。図3の場合は、請求項8は内的付加の引用形式請求項であり、請求項7を引用している。請求項9,10は外的付加の引用形式請求項であり、請求項9は請求項7,8を引用しており、請求項10は請求項7,8,9を引用していることを表している。

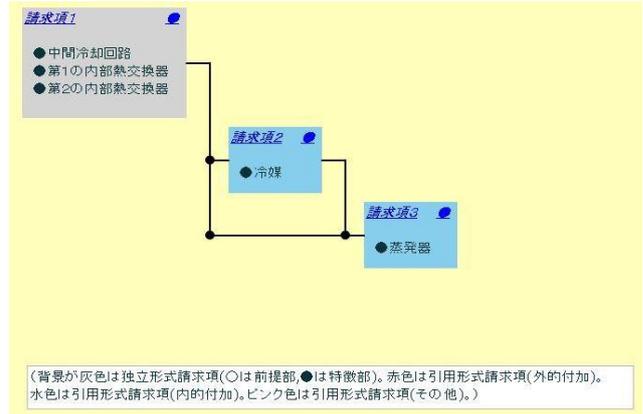


図2：「注釈付きクレームツリー」の例  
(特開 2004-101107 を表示)

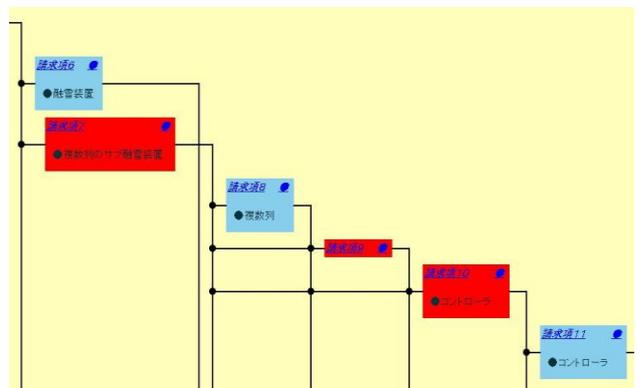


図3：「注釈付きクレームツリー」の例(一部)  
(特開 2004-176489 を表示)

## 2.3 特許請求項の細粒度解析

特許請求項は一文かつ長文で記述されることが多く、構文構造が複雑になってしまう。その上、独特の言い回しやスタイルで記述されるので、読解は非常に困難である。我々は請求項の構造解析を行うツールを既に開発済みであるが<sup>[6][7]</sup>、これらはトップレベルでの解析だけを行うものであった。「特許請求項の細粒度解析」は、より細かいレベルでの解析を行うものである。

まず、特許請求項を言語解析し、構成要素に分割する。その要素はノードとして表現する。そして、この構成要素及びその説明との間における関係を枝で表すことで、ツリー構造で出力する(図4)。ここで、出力する結果は、構成要素とその説明全て、構成要素の説明の要約、構成要素名のみ、の3種類の図を出力できるようになっている。このことによって、状況に応じた特許請求項の細粒度解析結果を得ることができる。

具体的に説明すると、構成要素とその説明全ては、構成要素及びその説明を省略しないで出力したものである。構成要素の説明の要約は、構成要素の説明が長くなる場合、先頭と末尾の数文節だけを表示するというものである。構成要素のみは、構成要素の説明を全て省いたものである。

図4は構成要素とその説明全てを表示した例である。「中間冷却回路」と「第1の内部熱交換器」と「第2の内部熱交換器」が構成要素として現れており、それぞれの要素の説明が何であるのか、見て分かるようになっており、確認できる。

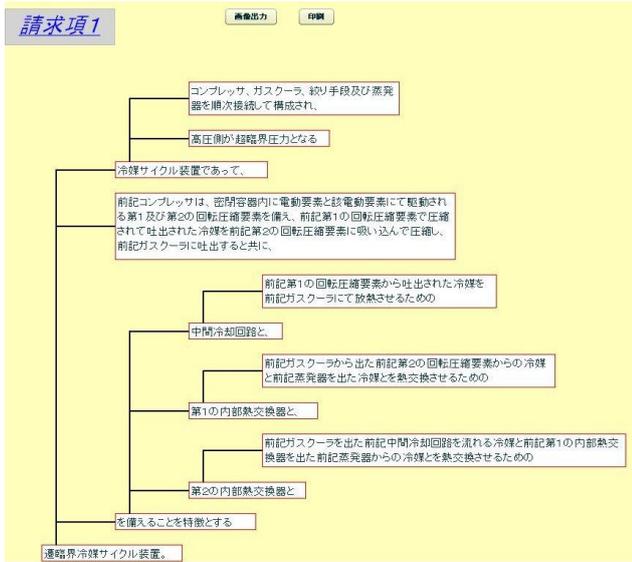


図4：「特許請求項の細粒度解析」の例  
(特開 2004-101107 請求項 1 を表示)

### 2.4 特許書類のハイパーテキスト化

特許書類は記述量が多く、例えば特許請求の範囲において、自分が求める請求項を探し出すのは手間がかかる作業になる。そこで、特許書類本文と「注釈付きクレームツリー」と「特許請求項構造解析」の関連付けを行うために、特許書類のハイパーテキスト化を試みる。具体的には、特許書類読解支援システムにおいて、「注釈付きクレームツリー」と「特許請求項の細粒度解析」の請求項にリンクを設けることで、特許書類の複数個所に記述されている内容と相互参照することができるようにし、その読解を支援する。

2.1 節において説明したように、特許書類本文は HTML で出力されており、「注釈付きクレームツリー」と「特許請求項の細粒度解析」は Flash で出力されている。そのためスクリプト言語間の連携、具体的には HTML の JavaScript と Flash の ActionScript 間の連携を行うことで、ハイパーテキスト化を実現する。

具体的には、特許書類の請求項記載部分にアンカーを設けて、「注釈付きクレームツリー」と「特許請求項の細粒度解析」の請求項部分をクリックしたとき、特許書類を出力するウィンドウ内において、そのアンカー先にジャンプできるようにしている。HTML でも請求項の切り替えボタンを用意しており、これを切り替えることで、Flash 内の「特許請求項の細粒度解析」で表示する請求項を切り替えるだけでなく、特許書類での該当する請求項へのジャンプも行うように連携されている。

また、特許請求項と「発明の詳細な説明」との間で、用語や記述内容の対応付けを行う。具体的には、「注釈付きクレームツリー」の各請求項表示ボックスに、「実施例へのジャンプボタン」を作成した。この「実施例へのジャンプボタン」を押すと、左側の「特許書類サブウィンドウ」において、当該請求項に対応する実施例が表示されるようにスクロールさせる。

### 3. 評価と考察

特許請求項を細粒度解析することの効果としては、トップレベルに存在する構成要素を明確化できることと、より深いレベルに存在する構成要素を検出し明確化できることが挙げられる。

図4における請求項例を挙げると、従来の解析ツールでは、図5で示したように「中間冷却回路」と「第1の内部熱交換器」と「第2の内部熱交換器」が構成要素として明確化されていなかった。しかし、「特許請求項の細粒度解析」を行うことで、図4で示す通り、これらの要素が明確化されるようになり、より理解しやすくなっていることが確認できる。

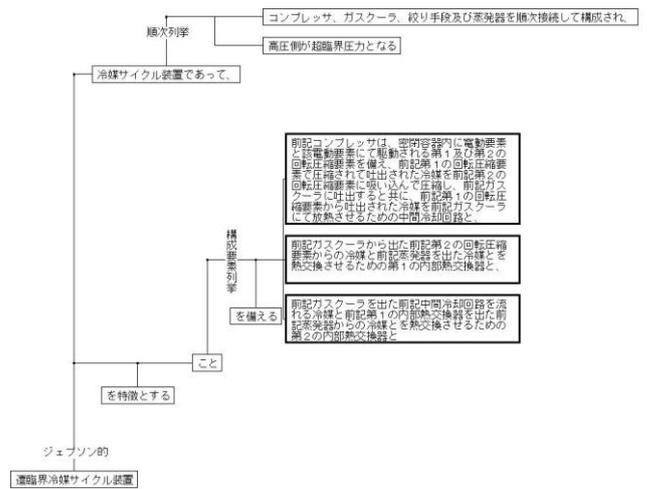


図5：従来の解析ツールでの結果表示の例  
(特開 2004-101107 請求項 1 を表示)

次に、図6における請求項例の場合は、「ダクト」「送風手段」「清浄化手段」の構成要素がトップより深いレベルにあるため、従来の検出ツールでは、検出されていないことが確認できる。しかし、「特許請求項の細粒度解析」を行うことで、図7に示す通り、これらの構成要素が検出されるようになった。

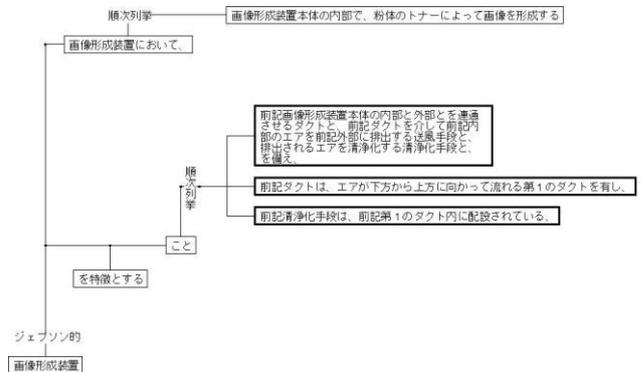


図6：従来の解析ツールでの結果表示の例  
(特開 2004-45573 請求項 1 を表示)

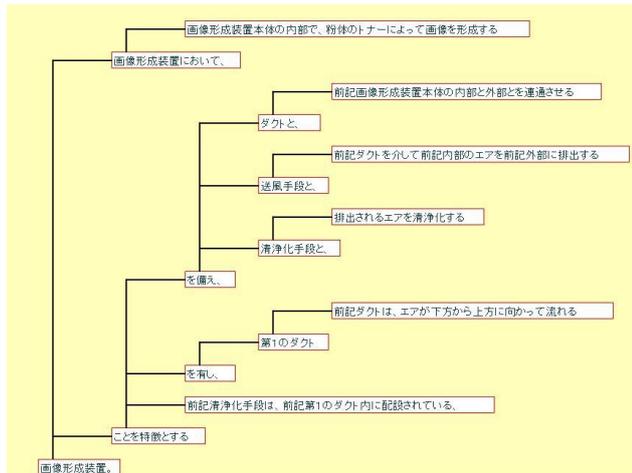


図7:「特許請求項の細粒度解析」の結果表示の例  
(特開 2004-45573 請求項 1 を表示)

それに加え、「特許請求項の細粒度解析」で得られた構成要素や他請求項の引用等の情報を「注釈付きクレームツリー」に付加することによって、多くの請求項があるような特許書類の場合でも、それぞれの請求項の引用関係が一目で分かるようになり、より読解支援に役立つ表示を行うことができるものとする。

また、「特許書類のハイパーテキスト化」をシステムに付加したことで、膨大な量になりがちな特許書類と、上記の「注釈付きクレームツリー」「特許請求項の細粒度解析」の対応付けを行うことができた。特許書類をブラウザで表示させようとする、一画面で表示できる情報は少なくなってしまうので、目的の箇所を見つけるにはスクロールを多用することになり、大変手間がかかる。そういう意味で、この特許書類読解支援システムの機能として、ハイパーテキスト化を付加することは、特許書類の読解を支援する上で有用な機能であると言える。

更に我々は、村田真樹氏が開発した、「請求項特徴的単語の強調表示機能」<sup>[8]</sup>を特許書類読解支援システムに取り入れた。具体的に説明すると、特許書類における特許請求の範囲を上から順番に探索し、特徴的な単語が現れたときに、その単語を請求項毎に定めた色で強調表示する。そして、それ以降で同じ単語が現れた場合、その単語を同じ色で強調表示する機能である。これにより、各請求項の特徴的な単語が一目で分かるようになり、それぞれの請求項の違いを容易に把握することができるようになる。また、発明の詳細部分でも同じように単語を強調表示することで、そこで現れた単語がどの請求項の特徴的な単語なのかを容易に把握できるようになる。この機能を取り入れることで、特許書類読解支援システムは、より読解支援する上で有用なものとなったと考える。

#### 4. 今後の課題とまとめ

本研究で作成した特許書類読解支援システムの根幹となっているものは、言語処理の部分であると言える。一般に自然言語は曖昧性を持つため、特に長文の構造解析は困難である。本システムでは、特許書類で多用される言い回し・表現に着目し、それを「手がかり句」として利用することで、構造解析を行っている。しかし、数学や化学分野のように、数式や化学式を含むものや、極めて長い物質名

を含む特許文書については正しくない解析をする場合もある。

今後の課題としては、こうした分野の特許書類の解析に対応できるように、言語処理部分の改良をすることが考えられる。また、特許書類読解支援システムにおいては、ハイパーテキスト化の部分で、「注釈付きクレームツリー表示機能」と「特許請求項の細粒度解析機能」の構成要素にもリンクを設け、特許書類の該当部分にジャンプできる機能を検討中である。

本研究は、特許に関する工学的アプローチである「特許工学」<sup>[9]</sup>の確立に向けた活動の一環である。特許ライフサイクルを支援するツール群における「分析・評価系ツール」<sup>[10]</sup>としての特許書類解析ツールの整備に努めていきたい。

#### 謝辞

本論文は、独立行政法人情報通信研究機構(NICT)の民間基盤技術研究促進制度に基づく委託研究「知的財産(特許・商標)構築・活用のための情報通信基盤技術の研究開発」の研究成果に基づきます。

#### 参考文献

- [1] 谷川英和, 新森昭宏, 渡辺俊規, “特許記述言語(PML)と統合的特許工学システム”, 第7回日本知財学会学術研究発表会 (2009).
- [2] 新森昭宏, 大屋由香里, 谷川英和, “請求項解析ツールを用いた請求項引用関係の分析”, 第6回日本知財学会学術研究発表会 (2008).
- [3] 新森昭宏, 大屋由香里, 谷川英和, “特許請求項における多重多数項引用の検出と書き換え”, 情報処理学会論文誌, Vol.49, No.7, pp.2692-2702 (2008).
- [4] 新森昭宏, 渡辺俊規, 谷川英和, “引用形式請求項における内的付加と外的付加の判定とタグ付けについて”, 第7回日本知財学会学術研究発表会 (2009).
- [5] 特許庁: 特許・実用新案審査基準 (2007). [http://www.jpo.go.jp/shiryoku/kijun/kijun2/tukujitu\\_kijun.htm](http://www.jpo.go.jp/shiryoku/kijun/kijun2/tukujitu_kijun.htm)
- [6] 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, Vol.45, No.3, pp.891-905 (2004).
- [7] 谷川英和, 新森昭宏, 大屋由香里, 奥村学, 難波英嗣, “特許明細書の作成・分析・評価を支援する統合的特許工学支援システム”, 第5回日本知財学会学術研究発表会 (2007).
- [8] 村田真樹, 金丸敏幸, 白土保, 馬青, 井佐原均, “種々の重要表現強調表示ツールバーの開発”, 言語処理学会年次大会発表論文集 (2007).
- [9] 谷川英和, 河本欣司, “特許工学入門”, 中央経済社 (2003).
- [10] 谷川英和, 森本悟道, “特許工学: 特許ライフサイクルに情報学を適用した新しい研究領域”, 情報処理, Vol.49, No.4, April (2008).