

# 特許文書の動作表現分析・学習に基づく発明上位化・下位化

下野 普也<sup>†</sup> 山本 岳洋<sup>†</sup> 谷川 英和<sup>††</sup> 山田 篤<sup>†††</sup> 宮脇 正晴<sup>††††</sup>

田中 克己<sup>†</sup>

<sup>†</sup> 京都大学大学院情報学研究科 〒606-8501 京都府京都市左京区吉田本町

<sup>††</sup> IRD 国際特許事務所 〒540-0008 大阪市中央区大手前1丁目7番31号 OMMビル8F 私書箱53号

<sup>†††</sup> 京都高度技術研究所 〒600-8813 京都市下京区中堂寺南町134番地

<sup>††††</sup> 立命館大学大学院法学研究科 〒603-8577 京都府京都市北区等持院北町56-1

E-mail: †{shimono,tyamamot,tanaka}@dl.kuis.kyoto-u.ac.jp, ††htanigawa@ird-pat.com,

††yamada@astem.or.jp, †††mmt23360@law.ritsumei.ac.jp

あらまし 本研究は、価値ある特許を作成するための発明上位化・下位化に取り組む。特許出願において、特許請求の範囲を定める請求項は極めて重要である。価値ある発明を特許文書として表現するためには、従来技術と被らなかつた範囲の権利を記すとともに、その権利範囲の下位概念も網羅的に記述することが重要である。そこで、本論文では、請求項において重要な役割を果たす動作表現に着目し、動作表現の上位化・下位化を行う手法を提案する。この技術を実現することで、ある特許文書中の請求項の上位化・下位化支援が行うことができると考えられる。従来技術において、名詞の上位語・下位語への置き換えの技術は存在するが、動作を上位化・下位化の技術は存在しない。例えば「冷蔵庫」と「電気機器」という語の上位・下位関係を獲得することは困難ではないが、「扉の開閉を検知すること」と「人を検知すること」という2つの動作の上位・下位関係を獲得することは難しい。本論文では、特許文書中の請求項の関係を分析することで、上位・下位関係にある動作表現を獲得する手法を提案する。具体的には、内的付加と呼ばれる関係にある請求項間や、ジェブソン形式と呼ばれる記述方式に従っている請求項から、上位・下位関係にある動作表現を獲得する。

キーワード

## 1. はじめに

特許申請に必要な書類は、「願書」「特許請求の範囲」「明細書」「要約」「図面」の5つで構成されており、その中で最も重要なものが特許文書の利権範囲を決定する「特許請求の範囲」である。特許請求の範囲は請求項という箇条書きの羅列によって記述され、請求項には独立請求項と呼ばれる他の請求項を引用していない請求項と、他の請求項を引用して記述される従属請求項の2種類が存在する。この独立請求項と従属請求項を組み合わせることで、特許請求の範囲が取得すべき複雑な利権の範囲を、階層的に表現することを可能としている。価値ある特許文書において特許請求の範囲に求められる性質は、

- 独立請求項において、従来技術と被らなかつた最も広範囲の利権をカバーすること
- 従属請求項において、その下位概念をくまなく記述すること

の2点である。第一項を満たすことが特許文書の性質上最も重要なのは当然であるが、第二項を満たす性質も重要である。例えば、独立請求項だけの特許請求の範囲を提出した場合、申請が通ればもちろん問題は無いが、もし申請が拒絶された場合、その独立請求項に対して新規性または進歩性が無いということしか判断が付かない。しかし、その下位概念をくまなく記述していれば、各請求項に対しての判定が返ってくるため、再申請

の際に請求項の修正がしやすくなる。また申請した特許に対して特許権侵害であると申し立てがあった場合、独立請求項だけで申請していれば、特許全体が無効となってしまうが、従属請求項をしっかり記述していると、その一部が無効になるだけで済むことがある。したがって、この2点を満たす特許請求の範囲を記述することが特許申請において重要である。

そこで、より価値のある特許請求の範囲、ひいては請求項を記述することを支援する技術とは、発明を自動的に上位概念化・下位概念化が出来ることである、と考えた。上位・下位概念の一般的な例として挙げられるのは、例えば「冷蔵庫」と「電気機器」といった関係である。こういった名詞の上位・下位関係を抽出する技術は元来存在し、特に珍しいものではない。しかし、発明において、例えば図1のように「扉の開閉により広告が切り替わる冷蔵庫」の上位概念には「人を検知することで広告が切り替わる冷蔵庫」が存在する。この例において、「扉を開閉すること」と「人を検知すること」は上位・下位関係であると言えるが、こういった動作や機能を表すフレーズの上位・下位関係を取得するような研究は存在しない。発明を自動的に上位概念化・下位概念化する技術を生み出すためには、こういった動作を表現するフレーズの上位・下位関係の知識を集めることが極めて重要であると感じた我々は、この関係の知識を特許請求の範囲の中から抽出することを考えた。

特許請求の範囲は前述の通り、独立請求項と従属請求項に

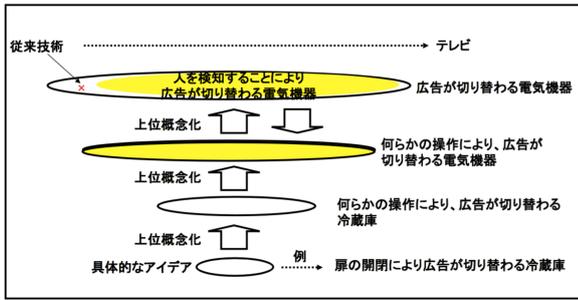


図 1 発明の上位概念化

よって階層的に記述されている。その中でも特に内的付加と呼ばれる関係は、厳密な意味で下位概念を記述されている場合が多い。また 1 つの請求項の中でも、ジェブソン形式と呼ばれる形式で記述された請求項は、従来技術を前提として新規技術を新たに書き足すという記述方式を行っているため、1 つの請求項の中で内的付加のような関係が存在する場合がある。こういった請求項の上位・下位関係の中には、その根幹を為す動作表現のフレーズがその関係を決定しているものが存在し、そのフレーズ対がそのまま抽出すべき知識である動作表現の関係であると考えた。以降の章ではこの動作表現対を抽出するための手法について説明する。

## 2. 関連研究

特許文書の構成や文章はわかりづらく、これらを対象とした研究は多く存在する。新森らは特許文書の可読性を高めるために、手がかり句という特許文書を書く上で一つの文法を利用した構造解析手法を提案した [1]。高木らはさらに、細粒化された構造解析を可能にし、また請求項の引用関係を可視化することにも成功している [2] [3] [4]。こういった構造解析手段は、PML(特許記述言語) と呼称されるマークアップ言語の開発にも役立つ [5]。PML とは、元々特許文書には特許庁が指定するフォーマットの XML が存在するのだが、これでは高度なツールを利用するさいに構造の粒度が大きい等の問題が存在するため、新たに開発されたマークアップ言語である。現在では特許文書を自動的に PML へと変換する手法も確立されており本研究ではそのツールを利用する。

請求項間の関係に着目した研究も存在する。引用形式請求項において、複数の請求項を引用することを「多数項引用」と呼び、さらに引用元が多数項引用となっていることを「多重多数項引用」と呼ぶが、多重多数項引用を用いた特許の記述はアメリカ等海外では認められていない事が多く、これを自動的に書き換える手法を新森らは提案している [6]。

また、請求項で記述される用語は極めて特殊なことが多く、可読性や互換性に乏しいと言える。そこで、明細書中に存在する対応する文章と照らし合わせて可読性を高める手法や [7] 特許文書を引用する論文が存在する場合、その引用関係を利用して論文用語と特許用語の変換を行う手法が提案されている [8] [9]。

## 3. 概念説明

本節では特許文書特有の概念について説明する。

### 3.1 請求項

特許請求の範囲において、記述される保護を受けたい発明を記載した項である。請求項は複数記載することができ、それぞれの請求項に記載された発明についてそれぞれ特許権の効力がある。さらに請求項は以下の 2 つに分類することが出来る。

- 独立請求項  
他の請求項を引用しないで記述する形式
- 従属請求項  
他の請求項を引用して記述する形式。引用方法については 3.2 節にて説明する。

一般的に独立請求項で上位概念を記述し、従属請求項でその下位概念を記述する。

### 3.2 内的付加と外的付加

内的付加と外的付加は従属請求項において、他の請求項を引用する手法である。神山ら [10] によれば、それぞれ次のように定義される。

- 内的付加  
内的付加は、上位概念に内包される下位概念を案出するものであり、それには限定、選択等が含まれる。内的付加には、発明それ全体についての内的付加と、発明を構成する各要素についての内的付加とがある。
- 外的付加  
外的付加は、現状の構成要素の組み合わせに対して別の構成要素を付け加えるものである。その結果、それら全体として新しい技術的価値が生じるのであれば、それは発明を拡充する展開であると同時に、別発明の案出でもある。

### 3.3 ジェブソン形式

「～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「名詞まとまり」とそれに対する修飾部が前置されている形式である。特に、後半部の修飾部は「～を特徴とする～」という形で記述されることが多い。前半部では、公知部分(すで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。

### 3.4 構成要素

構成要素とは発明を成立させるために存在する複数のパーツである。ほとんどの場合、発明は、複数の構成要素と各構成要素間の関係だけで成立している。例えば冷蔵庫の場合であれば、「扉」「冷却装置」「照明」などの構成要素が存在し、その組み合わせで表現される発明である。構成要素は「扉」のような一般的な語であっても、その特許文書中では固有名詞としてそれ以上の限定力を持つ。

## 4. 提案手法

請求項の内的付加の関係を用いて、機能表現フレーズを抽出することを考える。ここで、ジェブソン形式を用いて請求項において、その前半部と後半部で、特許の引用関係とほぼ同義の

関係があることから、ジェブソン形式を用いて記述された請求項の場合、その請求項内においても後述するアルゴリズムを適用することとする。

#### 4.1 請求項対の抽出

まず、機能限定の内的付加関係にある請求項のペアを取得する手法を提案する。特許文書の特徴を考慮した結果、言語パターンを用いることが効果的だと判明した。

手法は以下の通りである。x はその特許文書中の構成要素とする。

- (1) 「前記 x は、」という記述を抽出する
- (2) それより以前の記述で「～する(動詞の終止形)+x」という記述を抽出する

上記の手法を図2の例に適用させると次のような手順を踏む。

- (1) 「前記制御回路は、」という記述を抽出する(請求項3)
- (2) それより以前の記述で「～の制御を行う制御回路」という記述を抽出する(請求項2)
- (3) 以上の手順により、請求項3は請求項2の制御回路の機能を限定していると判定する

【請求項1】  
対向配置された一対の基板と、  
該一対の基板間に挟持された表示媒体と、  
該一対の基板のうち一方に形成された共通電極と、  
該共通電極を駆動する共通電極駆動回路と、  
前記共通電極と前記共通電極駆動回路との間を接続する配線と、  
前記共通電極駆動回路の電源供給側に設けられ前記共通電極に流れる電流を検出する検出部とを有する表示装置。  
【請求項2】  
さらに、前記検出部において検出された電流に基づいて表示に関する制御を行う制御回路を有する請求項1に記載の表示装置。  
【請求項3】  
前記制御回路は、前記表示装置を駆動する駆動回路の制御信号入力端子、または前記表示装置に入力されたビデオ信号に対する信号処理回路の制御信号入力端子に入力され、前記表示装置のコントラストとブライトネスとカラーサチュレーションとガンマ補正値とのうちから選択される少なくとも1の特性を制御する回路である請求項2に記載の表示装置。

図2 請求項の例1

それぞれの手順について説明を行う。「前記 x は、」の表現から始まる場合、図3のように、x の述部は「y する」という一般動詞である表現の場合が多い。

【請求項2】  
さらに、前記検出部において検出された電流に基づいて表示に関する制御を行う制御回路を有する請求項1に記載の表示装置。  
【請求項3】  
前記制御回路は、前記表示装置を駆動する駆動回路の制御信号入力端子、または前記表示装置に入力されたビデオ信号に対する信号処理回路の制御信号入力端子に入力され、前記表示装置のコントラストとブライトネスとカラーサチュレーションとガンマ補正値とのうちから選択される少なくとも1の特性を制御する回路である請求項2に記載の表示装置。

制御回路の機能の説明

制御回路の機能の限定

図3 「前記 x は、」

一方で「前記 x が」など他の表現により x が限定される場合、「z である」など、x の状態を限定する場合はほとんどであった。

さらに、限定される前の x について、その機能を説明してものでなければ、機能表現の上位・下位を取得することが出来ない。そこで x が特許で初めて出現した場合において、その機能について表現されているときの共通点を模索すると、直前が動詞の終止形がほとんどであることを確認した。図4のように、また、x がそもそも機能について表現されている場合、その限定も機能についての限定であることが多いため、両手法を併せることにより、より良い精度で機能限定の内的付加関係にある

【請求項1】  
光取出しを向上させる構造体を有する発光ダイオード(LED)であって、  
エピタキシャル成長させたp型層と、エピタキシャル成長させたn型層と、前記p型層および前記n型層の間のエピタキシャル成長させた活性層とを備えた発光ダイオード構造体と、  
該発光ダイオード構造体の下に隣接するスペルゲド層と、  
前記発光ダイオードと一体化された光取出し構造体を備え、前記光取出し構造体が、前記発光ダイオード内に捕捉された光が反射および/または屈折して前記発光ダイオードから出て行くことを可能にする表面を備え、  
前記光取出し構造体は、0.1 nmから1 μmのサイズの粒子を含む分散体層を含むことを特徴とする発光ダイオード。

直前が「……する」という形ではない

図4 「前記 x は、」

請求項のペアを取得する事が可能となった。

#### 4.2 フレーズ抽出の前処理

次に、抽出した請求項から機能表現を表すフレーズを抽出することを考える。ここでは新森らの構造解析手段[1]を用いる。その手順は以下の通りである

- (1) 形態素解析  
字句解析の前処理として、単語の切り出しを行い、かつその品詞情報を得る。新森らは形態素解析には茶釜を使用している。
- (2) 字句解析  
形態素解析の出力結果を入力し、文脈を判定しながら、トークンと文字列のペアを出力する。トークンは図2.に示すものに、POSTP\_NO(名詞まとまり)中の助詞「の」に対して割り当てるトークン)とWORD(他の9種類のトークンが割り当てられなかった単語に対して割り当てるトークン)の10種類を割り当てる。
- (3) 構造解析  
生成したトークンを元に構造解析を行う。

手がかり句	割り当てるトークン
において(、ト)	JEPSON.CUE
に於(以)て(、ト)	
であって(、ト)	
にあたり(、ト)	
に当(た)り(、ト)	
を特徴と(した)する(、ト)?	FEATURE.CUE
を併せて構成され(た)る(、ト)?	COMPOSE.CUE
を(、ト)?(共)に備(そ)な(た)る(、ト)?	
を(、ト)?併置(した)する(、ト)?	
(で)から構成され(た)る(、ト)?	
を(、ト)?併置(する)した(、ト)?	
を(、ト)?併置(せ)た(、ト)?	
(から)より(、ト)?(なる)った(、ト)?	
(から)より(、ト)?(成)った(、ト)?	
を(、ト)?取(た)る(、ト)?	
(COMPOSE.CUEに対応する手がかり句のみに存在する)	(「(名詞 記号)」に対して) NOUN
(名詞 記号)と(、ト)	(「と」に対して) POSTP_TO
	(「(、ト)」に対して) PUNCT_TOUTEN
(「(動詞基本形 助動詞基本形)<名詞まとまり>」の前に存在する)	(「(動詞連用形 助動詞連用形)」または
(動詞連用形 助動詞連用形)(、ト)	(「(動詞基本形 助動詞基本形)」に対して)
(動詞基本形 助動詞基本形)と(共)とも)に	VERB.RENYOU
	(「(、ト)」または「と(共)とも)」に対して)
	PUNCT.TOUTEN
	(<名詞まとまり>に前置する「(動詞基本形 助動詞基本形)」に対して)
	VERB.KIHON

図5 特許請求項解析のための手がかり句

このアルゴリズムを元に、図2の請求項2,3を構造解析すると図6のような結果が得られる。本論文中では、構造解析の結果分割された各要素、(図6中の枠で括られた各項目)をブロックと呼称することとする。

#### 4.3 フレーズ単位での抽出

構造解析した結果を元に、以下のアルゴリズムによってフレーズの抽出を行う。

- (1) 限定の対象となっている構成要素を x とする。もし x が複合名詞(制御回路等)の場合は末尾の名詞を x' とする。

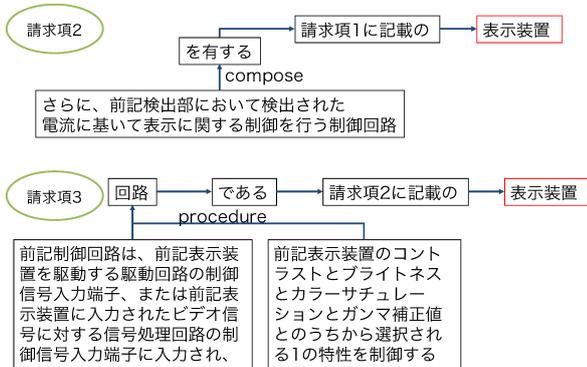


図 6 構造解析結果

- (2) 上位概念が含まれる請求項において、 $x$  または  $x'$  が含まれるブロックを抽出。同ブロック中に  $x$  または  $x'$  の直前に動詞の終止形が存在する場合、そのブロックを抽出する。存在しない場合、その直前のブロックも抽出する。もし、そのブロックに procedure タグが付いている場合はその並列のブロックも同様に抽出する。
- (3) 下位概念が含まれる請求項において、(2) と同様の手順を行う。ただし、「前記  $x$  は、 $\sim$ 」という文字列が存在するブロックより以前のブロックを抽出した場合、それを破棄する。最後に「前記  $x$  は、 $\sim$ 」という文字列を破棄する。
- (4) (3) で、該当する  $x$  も  $x'$  も見つからなかった場合、「 $\sim$ (動詞の終止形)+ことを特徴とする」という文字列を見付け、該当するブロックに対して (3) 同様の処理を行い、前記文字列を破棄する。

図 6 に上記アルゴリズムを適応させると以下のような手順を踏む。その結果は図 7 に示す通りになった。

- (1) 制御回路を  $x$ 、回路を  $x'$  とする。
- (2) 請求項 2 において、 $x$ (制御回路) が含まれるブロックを抽出する。
- (3) 請求項 3 において、 $x'$ (回路) が含まれるブロックと直前のブロックを抽出する。直前のブロックがその前と並列なのでさらに前のブロックも抽出する。

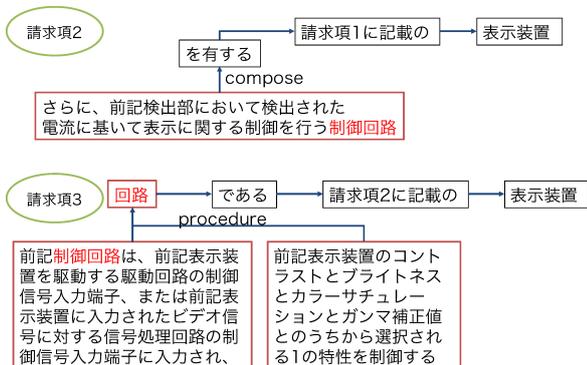


図 7 フレーズ抽出の工程

#### 4.4 実行結果

以上のアルゴリズムを手動で実行した結果、次に示すような結果が得られた

- 図 2 の特許請求の範囲からの抽出結果  
上位概念  
「さらに、前記検出部において検出された電流に基づいて表示に関する制御を行う」  
下位概念  
「前記表示装置を駆動する駆動回路の制御信号入力端子、または前記表示装置に入力されたビデオ信号に対する信号処理回路の制御信号入力端子に入力され、前記表示装置のコントラストとブライトネスとカラーサチュレーションとガンマ補正值とのうちから選択される 1 の特性を制御する」

- 図 8 の特許請求の範囲からの抽出結果  
上位概念  
「前記記憶媒体処理装置から会員記憶媒体の識別情報を受け付けた場合に、前記判定手段により重複登録と判定された複数の会員記憶媒体のいずれかが検知されたならば、重複登録された会員記憶媒体が存在する旨を外部へ出力する」  
下位概念  
「重複登録と判定された複数の会員記憶媒体の識別情報を前記記憶媒体処理装置に対して送信して、該複数の会員記憶媒体の識別情報の中から会員により利用される会員記憶媒体の識別情報を選択させるよう指示する」

【請求項 1】

遊技店に会員登録された会員が所持する会員記憶媒体の識別情報と該会員に関する情報とを対応づけて管理する会員管理装置と、該会員管理装置に接続され、少なくとも前記会員記憶媒体に係る処理を行う記憶媒体処理装置とを通信可能に接続した会員管理システムであって、

前記会員管理装置は、

遊技店に会員登録された会員が所持する会員記憶媒体の識別情報と該会員に関する情報とを対応づけて管理する管理手段と、

前記管理手段で管理される前記遊技客に関する情報の一緻度に基づいて、同一の情報が複数の会員記憶媒体の識別情報に対応付けられた重複登録が存在するか否かを判定する判定手段とを備え、

前記記憶媒体処理装置から会員記憶媒体の識別情報を受け付けた場合に、前記判定手段により重複登録と判定された複数の会員記憶媒体のいずれかが検知されたならば、重複登録された会員記憶媒体が存在する旨を外部へ出力する出力手段

を備えたことを特徴とする会員管理システム。

【請求項 2】

前記出力手段は、重複登録と判定された複数の会員記憶媒体の識別情報を前記記憶媒体処理装置に対して送信して、該複数の会員記憶媒体の識別情報の中から会員により利用される会員記憶媒体の識別情報を選択させるよう指示することを特徴とする請求項 1 に記載の会員管理システム。

図 8 請求項の例 2

## 5. まとめと今後の課題

本稿では、価値ある特許を作成するための発明上位化・下位化を目的とし、特許文書からの動作表現に関する上位概念・下位概念の抽出手法を提案した。提案手法は、請求項内、および請求項間の関係を用いることで、請求項から、上位・下位関係にある動作表現を含むフレーズを抽出する。今回は、少数の特許文書に対して、提案したアルゴリズムを手動で実行することで、上位・下位の関係にある動作表現を含むフレーズが抽出できることを確認した。一方、今回得られたフレーズは、粒度が粗く、1 章で述べたような「扉の開閉により広告が切り替わる」と「人を検知することで広告が切り替わる」といった上位・下位関係にそのままでは適用できないと考えられる。

今回取得出来たフレーズは大枠であり、必ずしも意味が 1 対 1 対応している部分の対応までは取れていない。その原因は、内的付加を行っている請求項で限定されている機能があくまで

必ず 1 対 1 対応しているわけではないことに起因する。この問題を解決するためにはより多くの特許文書の構造を解析し、その特徴を見極めなければならない。幸いにも、今後研究において、高木らの手法 [4] を自動的に実現するツールを研究目的で利用出来そうなので、特許文書の細かい構造から新たな特徴を発見出来るよう尽力したい。またこのツールのおかげで、手動でしか実現出来てなかった今回の提案手法の自動化は実現可能になると考えている。

今後は、この問題を解決し、また提案手法を自動化することで、システムが自動的に上位・下位の関係にある動作表現を抽出するシステムを実現する予定である。また、提案システムを用いて実験を行うことで、提案手法の有用性についても評価する予定である。

謝辞 本研究の一部は、JSPS 科研費 15H01718, 16H02906, 16K16156 によるものです。ここに記して謝意を表します。

#### 文 献

- [1] 新森昭宏, 奥村学, 丸川雄三, 岩山真ほか: 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).
- [2] 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).
- [3] Takagi, A. S. S.: 「特許請求の範囲」読解支援のための言語処理技術の改良と統合化.
- [4] 高木慎也, 新森昭宏: E-047 特許書類の可視化とハイパーテキスト化 (E 分野: 自然言語・音声・音楽, 一般論文), 情報科学技術フォーラム講演論文集, Vol. 9, No. 2, pp. 333-336 (2010).
- [5] 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).
- [6] 新森昭宏, 大屋由香里, 谷川英和ほか: 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).
- [7] 新森昭宏, 奥村学: 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, 自然言語処理, Vol. 12, No. 3, pp. 111-128 (2005).
- [8] 釜屋英昭, 難波英嗣, 相沢輝昭, 新森昭宏, 奥村学: 特許, 論文間の引用関係を用いた論文用語の特許用語への変換, 言語処理学会第 12 回年次大会, pp. 723-726 (2006).
- [9] 難波英嗣, 奥村学, 新森昭宏, 谷川英和, 鈴木泰山: 特許データベースからのシソーラスの自動構築, 言語処理学会第 13 回年次大会, pp. 1113-1116 (2007).
- [10] 神山公男: 明細書品質を向上させる発明者面談技法 (特集良い明細書の作成方法), パテント, Vol. 61, No. 7, pp. 30-47 (2008).