

請求項ブラウザの研究開発概要

Research and Development of Patent Claim Browser

松尾 衛*
Mamoru Matsuo

特許明細書には特有の表現があり、一般の技術者には記述内容を理解することが容易ではない。そこで、株式会社パトリス殿と共同で特許の請求項の可読性向上を目的とした研究を行った。本研究開発では、特許文書の中から請求項特有の言い回し等があり瞬時に内容を把握することが難しい【特許請求の範囲】に注目し【特許請求の範囲】の可読性向上に関する検討を行った。その結果を元に【特許請求の範囲】の文書構造に基づいて記載内容を整理して表示するプロトタイプシステム「請求項ブラウザ」を開発した。開発した請求項ブラウザの有効性について確認するため、特許出願の経験がある技術者を対象として実証実験を行った。本稿では「請求項ブラウザ」の概要と、有効性確認の評価実験の結果についてまとめた。

In this paper, we propose a browser system that supports understanding of Japanese Patent Claim. It is difficult to understand Japanese patent claims. We assume that a difficulty of Japanese patent claim is caused by a long sentence of claim and a complexity of claims relation. We solve this problem with the browser system, which analyzes structures of Japanese Patent Claims and arranges claims structure view. The browser system divides long sentences into proper size of phrases, and show divided phrases according to phrase placement rules.

The browser system is composed following technical factors.

- Phrase Dividing Rules
- Image and Caption Pairs Extraction
- Important Part Extraction
- Claim Relation Extraction
- Claim Relation Diagram Drawing

We show details of these technical factors, overview of the browsing system and evaluation of the browser system. The evaluation shows that the browser system is effective for understanding patent claims in short time.

1. まえがき

近年の知的財産に関する意識の高まりにより、社内において技術者自身が特許出願された先行技術を確認する機会が増加しつつある。しかし、特許明細書には特有の言い回しがあり、特許明細書を読むことに慣れていない人が読むと非常に読みにくいという印象を持つことがしばしば見受けられる。

そこで、本研究開発では特許明細書を読むことに慣れていない技術者を対象として、特許明細書の理解／利用を支援するための方法を検討した。特に特許明細書の請求項(【特許請求の範囲】)について理解支援を行うための方向性について検討を進めた。請求項には1文で発明内

容を記載する等の制約や、請求項特有の言い回し等があり、瞬時に内容を把握するのが難しい場合が多い。この問題に対し、以下の観点に基づいて、請求項の可読性向上を行うことを目標とし、請求項の内容を整理して表示するプロトタイプシステム「請求項ブラウザ」の研究開発を進めた。

- 請求項の文を、意味区切りとして適切な位置で区切ることにより、短文化をはかり可読性を向上させる。
- 請求項の従属関係を抽出し、独立請求項と従属請求項の役割の違いに基づいて、請求項の注目箇所を抽出する。
- 請求項の従属関係を視覚的に提示することにより、請求項間の全体構造把握を支援する。

なお、本研究開発は、2003年度の下期から2004年度

の上期にかけて株式会社パトリス殿と共同で開発を進めたものである。

2. 請求項ブラウザの概要

2.1 請求項ブラウザの実装方針

請求項ブラウザは研究者やエンジニア等の新たな顧客向けの特許文書内容整理サービスとして展開することを目標としている。研究者やエンジニアの場合、自分の専門分野であれば、ある程度のポイントを押さえるだけでも明細書の内容を把握することが可能と予想される。そのため、低コストで実現可能な、機械的な文書内容整理を実現するだけでも、利用者の明細書内容把握の支援が可能ではないかと考えられる。この新サービスの考え方を元に、特許文書の請求項を整理して表示することを目標に作成した。

本稿において、独立請求項及び従属請求項という用語が出現しているが、この用語は特許明細書の請求項の一般的な記載内容に関連する用語である。請求項は、図2.1-1の【請求項2】が請求項1を参照しているように、他の請求項を参照し、その請求項に要素を付け足すような記載が多く見られる。ここで、【請求項1】のように、他の請求項を参照していない請求項のことを独立請求項と呼び、【請求項2】のように他の請求項を参照している請求項のことを従属請求項と呼ぶこととする。

請求項ブラウザでは、以下の観点での構造化表示に重点を置いて開発を進めた。

- ・ 請求項の構造化表示(独立請求項/従属請求項別)
- ・ 請求項関連図の表示
- ・ 明細書本文の参照(現時点では未実装)

2.2 システム概要

請求項ブラウザの画面を図2.2-1に示す。図2.2-1に示す

<p>【請求項1】自然言語で記述した文章を検索要求文書として入力し、検索要求に類似した文書を検索結果として出力する情報検索装置であって、前記検索要求文書に含まれるキーワードを該キーワードの意味上の役割別に抽出するキーワード自動抽出部と、該キーワード自動抽出部により抽出されたキーワードの意味上の役割に基づいた検索条件式により検索対象文書を保存しているデータベースから前記検索要求文書に類似する文書を検索する検索部と、を有することを特徴とする情報検索装置。</p>	1文で記述
<p>【請求項1】は他の請求項に従属しない記述である。 →独立請求項</p>	
<p>【請求項2】前記検索部は、超並列計算機の複数の異なるプロセッサ上に複数の異なる検索条件式を設定し、前記検索対象文書を該複数の異なる検索条件式で同時並行的に全文検索し、前記複数の検索条件式に合致した結果を出力することを特徴とする請求項1に記載の情報検索装置。</p>	1文で記述
<p>【請求項2】は【請求項1】に従属した記述である。 →従属請求項</p>	

図2.1-1 請求項の例

ように、請求項ブラウザでは1文で記載されている請求項の文を意味区切りで分割し、分割後の短い意味区切りで表示している。また、文中から請求項の関係を自動抽出し、請求項同士の関係を請求項関連図にて図示することが可能である。図2.2-1に示した箇所の機能は以下のとおりである。

- (1) 独立請求項表示画面
- (2) 従属請求項表示画面
- (3) 請求項関連図の表示ボタン
- (4) 請求項関連図画面
- (5) 重要度が低いと思われるフレーズ(小さなフォントで表示される)

請求項ブラウザの独立請求項表示画面を図2.2-2に掲載する。図2.2-2に示した箇所の機能は以下のとおりである。

- (1) 複数回出現文字列
- (2) 欄の説明
- (3) 従属する請求項へのリンク
- (4) 画像の補足説明の表示ボタン

請求項ブラウザの従属請求項表示画面を図2.2-3に掲載する。図2.2-3に示した箇所の機能は以下のとおりである。

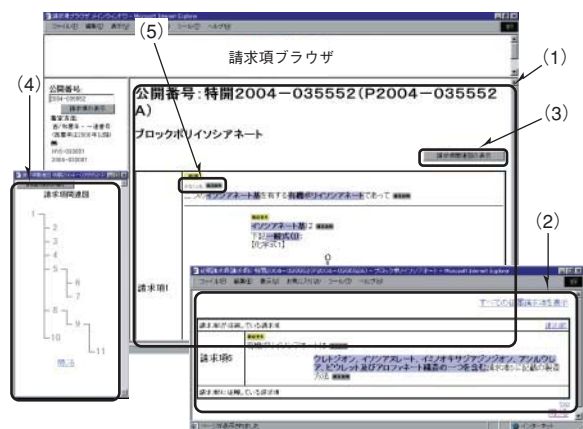


図2.2-1 請求項ブラウザ機能概要

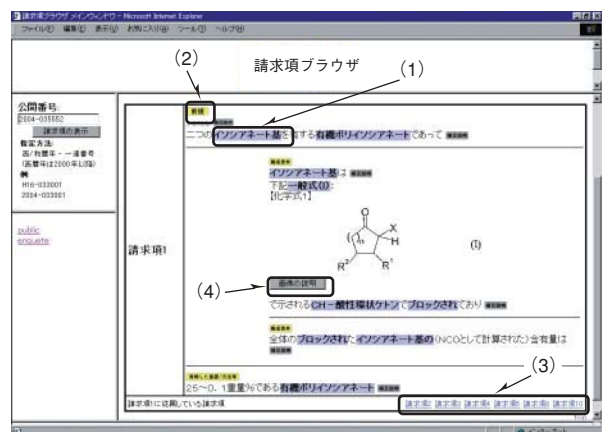


図2.2-2 請求項ブラウザ独立請求項表示機能概要

- (1) すべての従属請求項の表示リンク
- (2) 請求項が従属している請求項へのリンク
- (3) 欄の説明
- (4) 初出語句
- (5) 従属する請求項へのリンク
- (6) 画像の補足説明の表示ボタン

なお、図2.2-1～図2.2-3に示した請求項ブラウザの出力は、テキスト形式の公開公報もしくは公開公報DVD-ROMのXML形式で記載された公開公報から、自動的に作成することが可能である。ただし、画像を含む請求項の場合は、XML形式で記載されている必要がある。

請求項ブラウザの処理対象は自然言語であるため、現状の計算機による自然言語処理技術では、人間の判断と比較した場合、どうしてもある程度の誤りを含んだ出力となることがある。この誤った解析に対する対応の一つの考え方として、更に研究を進め、処理誤りに対する改善を進めた上でシステム化を行う方法が考えられる。しかし、この方針をとったとしても、完全に処理誤りを排除することは不可能であると考えられる。そのため、今後更に検討を進めても、その検討結果は「処理誤り何%改善」というような指標値が報告されるのみになってしまう場合も多くあり、システムの全体イメージは他の関係者に伝えられないまま、検討のみが進む恐れがある。

請求項ブラウザは、上記問題を考慮した上で、少しでも多くの関係者に我々研究開発者のシステムに対するイメージをつかんでいただくことを目的に作成を行った。そのため、人間の判断と比較して、ある程度の処理誤りを含む状態であっても、システム上での動作を最優先事項として実装した。

以上の理由により、請求項ブラウザの出力は(作成者の意図通りの処理がなされているにもかかわらず)出力結果が人間にとって非常に見にくい結果となってしまう

場合も存在する。しかし、そのような不適切な結果を含んでいる結果であったとしても、「処理誤り率何%」という値よりも直感的に、現状のシステム実現状況を伝えられるのではないかと考えている。

現状の請求項ブラウザの内部処理部分は、UNIX系のOS上に構築されるシステムであり、コマンドライン上から実行されるデータ入力部と、データ入力部の処理結果をHTMLに変換するブラウジング処理部から構成される。請求項のブラウジングはWebブラウザを使用して行われ、利用者はブラウジング処理部を意識することなく、ブラウジングすることが可能である。なお、推奨WebブラウザはMicrosoftのInternet Explorer 6である。ただし、現状の請求項ブラウザは、あくまで現状の技術的背景に基づいて開発したものであり、今後、利用者による評価結果に基づいたシステムの改善や、システムが提供可能な情報に基づくサービス提供方法の検討を進める必要がある。したがって、今後の評価/検討結果に基づいて、更に発展的な方向にシステムの最終実装イメージを固め改善を進めたいと考える。

3. 請求項ブラウザの構成技術

3.1 フレーズ抽出手法

図2.2-1～図2.2-3に示したように、請求項ブラウザでは、請求項の文を適切な意味区切りで文を分割して表示を行っている(以降、この意味区切りで分割した単位をフレーズと呼ぶ)。このように長い文の請求項に対し短文化を行うだけでも、読み手が一度に記憶する単位を減らすことができるため、請求項の文をフレーズに分割して切り出すだけでも、利用者の請求項読解支援へつながると予想している。

フレーズ分割を行うためには、フレーズをどのように抽出するかを決定する必要がある。本研究開発では、請求項に対する形態素解析結果を利用するフレーズ分割ルールに基づいてフレーズ抽出を行っている。なお、形態素解析は、奈良先端科学技術大学院大学の松本研究室により開発された形態素解析ツール茶筌を使用している。フレーズ分割ルールは、請求項の形態素解析結果の分析に基づいて作成されたものであり、意味区切りとして適切な箇所では請求項が区切られるように分析し、フレーズ分割ルールとして定義したものである。現在使用しているフレーズ区切りルールの抜粋を表3.1-1に掲載する。以上のルールにより入力した文書からのフレーズ抽出処理を自動的に行っている。

3.2 フレーズ中の図の扱い

請求項の文中には、文字のみでなく図(画像)を含む請求項がいくつか見受けられる。請求項に出現する画像は

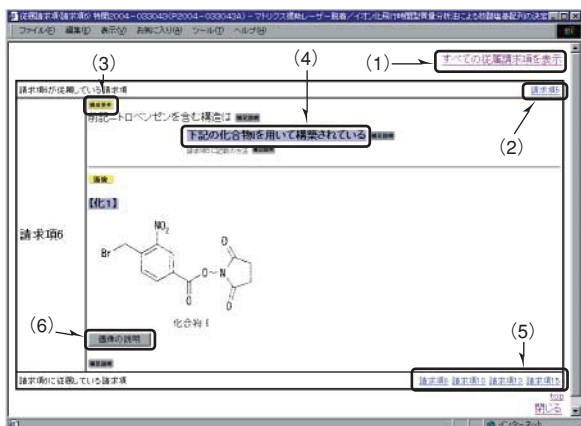


図2.2-3 請求項ブラウザ従属請求項表示機能概要

多くの場合、【図1】【化学式1】等の墨付き括弧の次に画像本体が出現し、場合によっては画像の後に画像の説明が出現する構造となっている。そのため、墨付き括弧から画像の説明までを1つの意味単位として扱うほうが、後の処理が簡潔になる。我々の手法では図3.2-1に示すように墨付き括弧から画像の説明までを1つの画像として扱う実装とした。以降この単位を画像単位と記す。この画像単位を画像毎に一意となるID文字列に変換し、3.1節に示したフレーズ分割処理を行う。出力時に画像単位を元に復元することにより、画像を含む文を通常の文と全く同様に処理できるような実装としている。

なお、画像単位は、画像の出現位置により抽出方法が異なる。画像の出現位置は大きく以下の3種類に分類することができ、画像がどの場所に出現しているかは、請求項中の句点の位置及び括弧の位置により自動判定を行っている。

- (1) 請求項の末尾に画像が出現する場合
- (2) 請求項の文中に画像が出現する場合
- (3) 画像の説明中に画像が出現する場合

3.3 重要な語句の抽出

2.1節において、特許文書の請求項は独立請求項と従属請求項が存在することを述べたが、従属請求項では、以下の理由により、多く出現する語句以外の語句が重要な語句である場合が多いことが確認された。

- (1) 従属請求項は、その請求項の従属元となる請求項で記載されなかった内容を補足するためにある請求項である。そのため、その請求項の従属元となる請求項で

表3.1-1 フレーズ分割ルール (一部抜粋)

No.	ルール内容
1	【】内はフレーズ抽出の対象外とする。
2	フレーズは名詞から開始する。ただし、非自立の名詞はフレーズ開始の対象外とする。
3	「ことを特徴とする」という表現は、フレーズに含まれないものとする。
4	句点の手前でフレーズを終了する。

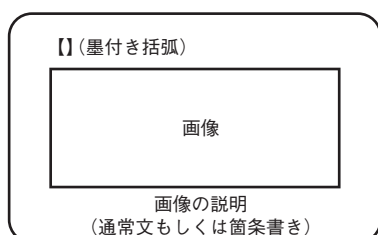


図3.2-1 画像として処理する単位 (画像単位)

出現していない内容(語句)が重要となる場合が多い。

- (2) 請求項では、権利を主張する内容を1度記載すれば十分であり、何度も記載する必要がない。そのため、その請求項に従属する請求項が存在しない従属請求項の場合は、1度しか出現しない語句が使用される場合が多い。

このように、従属請求項の場合は、複数回出現するという観点よりも、以前の請求項に出現せず、その従属請求項で初めて出現するという観点(以降、この観点を「初出」と呼ぶ)が重要なのではないかと考えられる。請求項ブラウザの従属請求項の表示では、この分析に基づいて、従属請求項にて初出する語句に注目し、その語句を重要部分としてハイライト表示を行っている。

3.4 請求項の従属関係の抽出

現在、請求項ブラウザにおいて請求項の従属関係の抽出に利用している表現と、その表現に基づいた現状における請求項関係抽出ルールを以下に示す。

- (1) 「請求項X～記載の」に代表される従属元参照表現全体を抽出する。
- (2) 項目(1)にて抽出された表現の間に出現する数字を、その請求項の従属元となる請求項番号として抽出する。ただし「Y項」(ただし、Yは数字)という記載の場合は請求項番号としての抽出は行わない。
- (3) 項目(2)で抽出した番号と、その前に出現した番号との間に、特定の表現がある場合は、抽出した2つの番号間の値を、その請求項の従属元となる請求項番号としてすべて抽出する。
- (4) 項目(1)に該当する表現が複数個ある場合は、項目(1)に戻り、次の表現を抽出する。すべての該当する表現に対して処理が完了したら終了する。

請求項の記載の仕方には、ある程度決まった書き方があるため、上記ルールでも比較的多くの請求項の関係を抽出することが可能である。しかし、人間が作成する以上、請求項の関係の記載にはさまざまな書き方がなされる可能性があり、今後もより多くの表現に対応可能なノウハウの蓄積が必要である。

3.5 請求項関連図の出力

3.4節に示した手法により抽出された請求項の従属関係を元に、以下のルールにより請求項関連図が作成される。

- (1) 横軸→独立請求項を第一階層(最上位)に配置し、従属請求項を順次第二階層以下へ配置する。ただし、従属元(親)請求項のうち、最も深い階層+1の階層に配置されるようにする。

(2) 縦軸→請求項番号の順に上から下に配置する。

(例)

「請求項1は独立請求項(従属元請求項なし)、請求項2の従属元請求項は請求項1、請求項3の従属元請求項は請求項2」の場合は図3.5-1のように配置される。また、請求項2と請求項3が、両方請求項1のみに従属し、並列の関係にある場合は、図3.5-2の請求項1～3の部分に示すような図となる。

上記により、配置した請求項番号に対して従属元から従属先のすべての関係について、結線を行うことにより、図3.5-3に示すような請求項関連図が作成される。

4. 評価実験

3章までに請求項ブラウザの方向性及び実装の概要について示したが、請求項ブラウザの有効性はあくまで仮定に基づいて検討されたのみである。そのため、請求項ブラウザの有効性をより客観的に確認するために、2004年の8月に三菱スペース・ソフトウェア社内にて有効性についての評価実験を行った。請求項ブラウザ評価実験の概要を本章に掲載する。



図3.5-1 請求項関連図における請求項の配置 その1

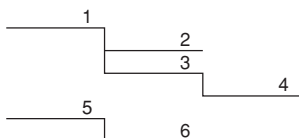


図3.5-2 請求項関連図における請求項の配置 その2

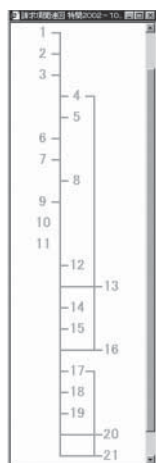


図3.5-3 請求項関連図

4.1 実験条件概要

実験は、社員13名を、請求項本文(プレーンテキスト)を用いて設問に解答する被験者(以降タイプAと記す)7名と、請求項ブラウザの出力を用いて設問に解答する被験者(以降タイプBと記す)6名に配分してWebによる設問に解答する形式で実施した。評価は、請求項ブラウザを用いたほうが、より短時間に解答内容の記載部分を請求項から見つけられるという仮定の下に、設問に対する回答時間で評価することとした。なお、被験者13名は特許出願経験のある技術者12名と知財担当者1名から構成され、タイプA及びタイプBへの配分は表4.1-1に示す配分となった。

表4.1-1に示した13名の被験者に対して、以下の手順により評価実験を行った。

(1) 評価実験の実施前に、あらかじめ被験者に評価実験の手順説明資料、紙に印刷した請求項(一部請求項関連図)を配布した。配布資料には以下の処置がなされている。

- ・評価実験の手順説明資料には、回答時間を計測することを明記し、実験の主旨に従い、設問に対する請求項内の記載を確認した上で回答するように指示する。
- ・紙に印刷した請求項は封筒に封入しておき、Web出題の回答直前に開封するよう指示する。
- ・上記紙に印刷した請求項は、タイプAの被験者には請求項をそのままプリントアウトしたものを送付し、タイプBの被験者には請求項ブラウザの出力をプリントアウトしたものを送付する。

(2) タイプBの被験者には項目(1)に示した配布資料の他に、請求項ブラウザの使用方法チュートリアルを配布する。更に請求項ブラウザを試使用することができるURL情報を送付し、評価実験実施期間前にあらかじめ請求項ブラウザを使用して使い方を覚えておくように指示をする。(ただし、チュートリアルで使用する請求項は、設問で使用する請求項とは別の請求項を利用した。)

(3) 評価実験実施期間に被験者はWebによる設問に解答する方式で評価実験を実施する。なお、評価実験実施期間は3日間に設定し、期間内であれば被験者はいつでも評価実験を行えるものとした。なお、回答時間はWebサーバにて自動的に記録される。

表4.1-1 被験者の配分

タイプA	タイプB
特許出願経験のある技術者 6名 知財担当者 1名	特許出願経験のある技術者 6名

4.2 評価結果

合計回答時間の平均を取った結果を表4.2-1に掲載する。さらに、回答者の回答時間のばらつきを押しやるため、タイプA、タイプBの各被験者のうち、最も回答時間が短かった被験者と、最も回答時間が長かった被験者の結果を除外して平均をとった結果を表4.2-2に掲載する。

表4.2-1及び表4.2-2に示すように、請求項ブラウザを使用したタイプBの評価実験を実施した被験者の方が、平均すると6～7分程度短い時間で設問に回答したことが確認できる。ただし、現状被験者の数も少ない状況であり、更に、回答時間のばらつきも大きな状況であるため、十分な評価結果であると断定できないことも事実である。とはいえ、多くが仮定の状態で進めてきた請求項ブラウザの有効性確認の第一歩としては、予想した方向での結果が得られ、検討を進める上での有効な要素を補強することができたのではないかと考えている。

5. むすび

本稿では、2003年度の下期から2004年度の上期にかけて株式会社パトリス殿との共同研究開発で進めた請求項ブラウザについて、その概要と以下の要素技術について示した。

- ・文の分割処理方法
- ・請求項の画像の扱い方
- ・重要部分の抽出方法
- ・請求項の従属関係の抽出
- ・請求項の従属関係の出力

また、請求項ブラウザの有効性評価のために、2004年の8月に三菱スペース・ソフトウェア社内にて実施し

表4.2-1 平均回答時間

タイプA平均回答時間	タイプB平均回答時間
36分14秒	29分21秒

表4.2-2 回答時間のばらつきを押しやえて求めた平均回答時間

タイプA平均回答時間	タイプB平均回答時間
34分34秒	28分40秒

た評価実験の概要とその結果を示した。評価実験の結果、請求項ブラウザを用いて設問に解答した被験者のほうが、25～40分程度の回答時間の設問に対し平均して6～7分程度短時間で回答したことが確認された。

6. 謝辞

本研究開発を共同作業としてすすめ、多くの協力をいただいた株式会社パトリス殿に感謝いたします。また、請求項ブラウザの評価実験にご協力いただいた社員の方々に感謝いたします。