

調査知識をルール化して利用した特許調査支援環境の提案

谷川 英和¹ 鶴原 稔也² 東 明洋² 鈴木 芳文² 中村 由紀子² 酒井 美里³ 渡辺 俊規¹ 増満 光¹

概要: 本稿では、特許調査実施工程を「検索式策定工程」と「選別工程」に大別し、これまで暗黙知であったものを知識ルールとして形式知化したことについて述べる。具体的には検索式策定工程では、「検索対象」「検索対象期間」「検索フィールド」「用語」「分類」について知識ルール化して利用した。また、選別工程において漏れている関連特許候補を提示したり、選別されているが非関連の特許候補を提示することができるようになる。

キーワード: 特許調査, 検索式提案, 調査知識のルール

Propose the supporting environment for patent search by using rules of search knowledge.

HIDEKAZU TANIGAWA¹ TOSHINARI TSURUHARA² AKIHIRO HIGASHI² YOSHIFUMI SUZUKI²
YUKIKO NAKAMURA² MISATO SAKAI³ TOSHINORI WATANABE¹ HIKARU MASUMITSU¹

Abstract: In this paper, a patent search implementation process is divided roughly into a "query decision process" and a "selection process", and it describes having made into the knowledge rules what is tacit knowledge until now, and having explicit knowledge. At the query decision process, the rules about the "search target", "search period", "search field", "term", and "classification" are constructed and used. And at the selection process, the another related patent candidates and non-related patent candidates among related patents are shown.

Keywords: Patent Search, propose of query, rules of search knowledge

1. はじめに

著者らは、2002年以降、発明の着想から権利化、権利行使に至る特許ライフサイクルにおける各作業について、工学的にアプローチを行う特許工学の研究を行っている[1]。特許工学は、特許ライフサイクルにおける各種作業に対して、方法論を抽出し、ツール（以下、「特許工学ツール」という）と教育により、方法論の普及を図ることにより、各種作業の品質と効率の向上を目指すものである。

また、特許ライフサイクルの上流工程を構成する特許調

査フェーズにおいて、その品質や効率は、調査者のスキルに大きく依存する。例えば、高度なスキルを有する特許調査者（以下、サーチャーという）は、適切な国際特許分類（IPC）、FI、またはFターム（以下、これらを総称して「分類」という）を利用し、かつ、一のキーワードとその同義語、上位概念語、下位概念語（以下、これらを「関連語」という）とを適切に利用し、漏れやノイズが少ない検索式を作成する。しかし、サーチャーでも、対応する技術分野の範囲が広い場合、適切な関連語を抽出することは容易ではない。また、特許調査において、特に、適切に分類を決定することは困難である。

このような課題を解決するために、調査者が入力した日本語のキーワードに対応する関連語と分類（IPC、FI、およびFターム）を提案し、選択された関連語と分類から検索式を

¹ 有限会社アイ・アール・ディー
IRD Corp.

² ドコモ・テクノロジー株式会社
DOCOMO Technology, Inc.

³ スマートワークス株式会社
SMART WORKS, Inc.

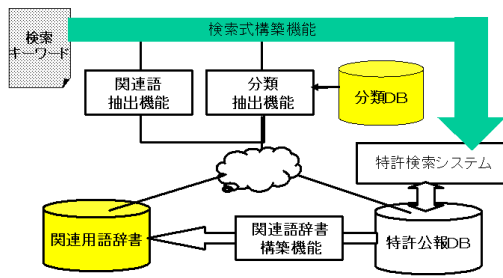


図 1 PSA の概要

Fig. 1 Outline of PSA.

構築する特許調査支援システム (PatentSearchAssistant, 以下、「PSA」という) を既に開発済みである [2]。PSA により、調査者は、キーワードを入力するだけで、その関連語やコードを含む適切な検索式を容易に構築することができるようになった。

しかし、現バージョンの PSA は、高度なスキルを有する特許調査者 (以下、サーチャーという) のノウハウの活用が不十分である。また、PSA を用いた特許調査でも漏れが生じ得る。また、特許調査の漏れの原因として、検索式の策定の不具合、および特許の選別の不具合の 2 通りの不具合が考えられる。なお、特許調査フェーズは、検索式策定工程、選別工程の 2 つのステップからなる。

さらに、検索式策定の不具合について、例えば、同一の意味の用語でも、企業により異なる用語を特許書類中に用いているが、他社の用語について知見がない場合である。また、特許選別の不具合について、例えば、非関連特許とした特許書類の中に関連特許が存在したり、関連特許と判断した特許書類の中に非関連特許が存在したりし得る。さらに、非関連特許とした特許書類、関連特許と判断した特許書類の情報を用いれば、関連特許の可能性のある特許書類を取得し得る。

そこで、著者らは、サーチャーのノウハウをルール化し活用できる環境、特許調査における漏れを極力防止できる機能を設計したので報告する。

2. PSA の概要

PSA は、キーワードを入力すれば、そのキーワードの関連語と IPC などの分類を用いた検索式を提案する。

PSA は、関連語辞書構築システムと検索式構築システムの 2 つのシステムで構成され、検索式構築システムで構築した検索式は既存の特許検索システムに渡される。このシステム全体の概要を図 1 に示す。

また、PSA における関連語の提案、および分類の提案の画面例を、図 2 に示す。図 2 は、調査者が入力したキーワード「データベース」および「検索」に対し、それぞれの関連語を提示していることを示す。また、図 2 において、調査者が入力したキーワードと調査者が選択した関連語 (以下、これらを「検索語」という) との関連性が高い



図 3 関連語抽出機能の出力例

Fig. 3 Output example of a related term extraction function.

分類を提案している。

2.1 関連語辞書構築機能

関連語辞書構築機能は、約 20 年分の特許公報 (公開特許公報、登録特許公報) から手がかり句を用いて用語間の関係性を取得し、関連語辞書を構築する [2]。本機能では、請求項における内的付加の従属項における限定的な用語を下位概念語、その被従属項の用語を上位概念語として抽出する。また、本機能では、例えば、「などの」と「等の」という表現の前後に記載されている用語は上位概念語、下位概念語の関係にあることに着目する。例えば、「HDD などの記録媒体」などである。この例では、「記録媒体」が「HDD」の上位概念語の関係にある。しかし、「などの」と「等の」だけでは、「パソコンなどのキーボード」のように上位概念語、下位概念語関係にない場合であっても登録を行ってしまう。そのため、「パソコンのキーボード」のように「などの」と「等の」を「の」に言い換えられる場合は、上位概念語、下位概念語の関係にないとした。また、「メモリや HDD などの記録媒体」のように共通の上位概念語、もしくは下位概念語を持つ用語が列記されている場合も上位概念語、下位概念語の関係にないとした。そして、共通の上位概念語と下位概念語を持つ用語は、同義語であるとした。

2.2 検索式構築のための機能

PSA は、調査者の入力したキーワードから関連語を抽出する関連語抽出機能と、キーワードと選択したキーワードの関連語 (以下、これらを「検索語」という) とから、適切な分類を抽出する分類抽出機能で構成される。

関連語抽出機能の出力例を図 3 に示す。図 3 において、調査者はキーワード「データベース」と「検索」を入力したとする。

関連語抽出機能は、キーワードに対応した関連語を関連語辞書から抽出する。抽出する関連語は、入力されたキーワードと、関連語抽出機能が出力した関連語のうちから調査者が選択した関連語である。



図 2 関連語と分類の提案

Fig. 2 Suggestion of a related term and a classification.



図 4 分類抽出機能の出力例

Fig. 4 Output example of a classification extraction function.



図 5 関連公報提示機能

Fig. 5 Presentation function of a Related Patent Gazette.

次に、分類抽出機能の出力例を図 4 に示す。

図 4 において、分類抽出機能は、関連語抽出機能で、調査者が選択した検索語を用いて、検索に適した分類を抽出する。抽出する分類は、すべての特許公報に対して、各特許公報に記載されている全用語数に対する検索語の出現割合を算出し、その特許公報が保持する各分類にその割合を重みとして与え、分類ごとに重みの総和（以下、「分類重要度」という）を算出し、分類重要度が高いものとした。そして、最終的に提案する検索式は、取得した分類の組み合わせと検索語（調査者入力のキーワードと、調査者選択の関連語）から構成される。この時、調査者は使用する分類を任意に選択し、検索式を組み替えることができる。なお、図 4 において、太字で表示されているコード「G06F 17/30」は、公報中の検索語の出現頻度が高く、PSA が重要であると判断したコードである。

2.3 分類の選択支援機能

調査者が選択した分類が正しいものであるか否かを判断するために、PSA は、抽出した分類の意味を出力する分類説明出力機能、および選択した分類が付されている特許公報を提示する関連公報提示機能を有する。

分類説明の出力例を図 4 に示す。分類説明出力機能では、各分類の説明を出力する。この説明は、調査者にとって理解し難い場合もある。そこで、検索語と選択した分類による検索され得る特許公報の中から N（例えば、10）の

特許公報を提示する関連公報提示機能も開発した。関連公報提示の画面例を図 5 に示す。

図 5 において、調査者は、任意の検索語を指定し、提案された分類のうち、「G06F 17/30」を選択したとする。すると、当該検索語と、分類「G06F 17/30」の両方を含む公報（発明の名称）が 10 件、表示される。また、各特許公報をクリックすると、公報の本文が表示される。

この時提示される特許公報は、同じような分類の組み合わせを可能な限り省いている。例えば、「G06F 17/30」と「G06F 17/60」の両方が付された公報が複数あった場合、その中の 1 つしか表示しない。また、分類が 2 つ付された特許公報と分類が 3 つ付された公報があった場合、付されている分類の少ない公報を優先的に表示する。これにより、より多くの分野の公報を提示することができる。また、調査者は、提示される複数の公報の中から、調査対象となる分野の公報を参考に、分類を選択することができる。

2.4 PSA についての検証結果

PSA が提案する検索式の有用性について検証した。本検証における特許検索は、特許電子図書館 (IPDL) [3] でを行い、用語の検索対象を「要約+請求の範囲」とした。また、本検証で調査対象として適切な公報（以下、「適合公報」という）であるかの確認は、検索結果から著者らが判断したものである。

本検証では、「OS」、「タスク」、「速度」をキーワードと

表 1 キーワードで構成された検索式の検索結果

Table 1 Search results of the search expression which consisted of keywords.

検索式	検索結果	適合数
OS * タスク * 速度	41件	13件

表 2 検索語とコードで構成された検索式の検索結果

Table 2 Search results of the search expression which consisted of a search word and a code.

検索式	検索結果	適合数
(OS+オペレーティングシステム+Operating System+基本ソフト+基本ソフトウェア) * (タスク+プロセス+ジョブ+スレッド+task) * (速度+スピード) * (G06F 9/46+G06F 9/45+G06F 9/50)	39件	35件

する調査を例に挙げる。3つのキーワードの積で構成された検索式で検索した検索結果と適合数とを表 1 に示す。

PSA が提案する検索式が、適合公報を絞り込むことに有用であるかについて、検索式を用いて検索した結果において、適合公報の割合が向上することを確認する。

本検証では、IPC を用いて行った。本システムが出力した複数の検索式から最も良い結果が得られた検索式の検索結果を表 2 に示す。

表 2 より、PSA が提案する検索式を用いることで、表 1 の結果に比べ検索結果に対する適合公報の割合を大幅に向上できたことがわかる。

3. サーチャーの知識をルール化した検索式チェック機能

検索式は、通常、「検索対象」「検索対象期間」「検索フィールド」「用語」「分類」を要素して有する。「検索対象」とは、検索する特許書類の国、および特許書類の種類を特定するものである。特許書類の種類とは、例えば、公開特許公報、特許公報、実用新案登録公報等である。また、「検索対象期間」は、検索したい特許の期間を指定するものである。期間とは、出願日、公開日、登録日等の期間である。また、「検索フィールド」は、検索式中の用語を検索するフィールドであり、例えば、「要約」「特許請求の範囲」「要約+特許請求の範囲」「全文」等である。「用語」は、キーワードである。さらに、「分類」は、日本国においては IPC, FI, または F タームであり、米国においては IPC または US クラスである。

PSA の次バージョンにおいて、検索式を構成する各要素について、サーチャーの知識ルールを用いて、妥当性をチェックする (図 6 参照)。また、特許調査において、特許調査の目的に応じて、要素の内容を変えるべきである。従って、知識ルールは、通常、調査目的毎に存在する。なお、調査目的とは特許調査の種類を言い、特許出願前に先行技術が存在しないか否かを調査する「出願前調査」、商品やサービスを市場に出す前に他社特許を侵害していないか否かを調査する「実施前調査」、他社の登録特許を無効にす

表 3 検索対象の知識ルール

Table 3 The knowledge rule for search.

調査目的	状況等	検索対象
出願前調査	特許出願前(a)※1	日本国公開系特許、実用新案
	特許出願前(b)※1	日本国公開系特許、実用新案 及び出願予定国の公開系公報 ※2 米国出願の場合は、登録系を考慮する ※3
実施前調査	特許成立分を調査	日本国登録系 特許、実用新案
	特許成立分+今後成立の見込みありを調査	日本国公開系及び登録系 特許、実用新案
無効資料調査		資料の種類・発行国を問わない。

※1出願前(a)は標準的な調査。出願前(b)は、日本出願の時点ではあるが、外国出願を行う予定の場合。
※2出願予定国によっては、特許/実用新案の種別が存在しないため、単に「公開系公報」とした。
※3米国の公開系公報発行は2001年3月～。近年のみ対象にした出願前調査であれば、公開系が良い。2000年以前に遡及したい場合は、登録系公報を検索対象に入れる。

表 4 検索対象期間の知識ルール

Table 4 The knowledge rule of a search period covered.

調査目的	検索期間
出願前調査	限定は不要
実施前調査	下記A、Bのいずれかで実施。なお両者を組み合わせてもよい。 (A) 実施予定国の権利期間内 (例：出願日＝過去20年分) (B) 検索期間の限定なし (別途、検索上で「法的状況＝権利生」に限定)
無効資料調査	調査対象特許の出願日、または優先権主張日に対し、 (A) 公知の範囲 (B) 先願+公知の範囲

表 5 検索フィールドの知識ルール

Table 5 The knowledge rule of the search field.

調査目的	検索フィールド
出願前調査	発明の名称+要約+特許請求の範囲
実施前調査	登録系であれば「特許請求の範囲」(+FI記号) 登録+公開系ならば「特許請求の範囲+全文」(+FI記号)
無効資料調査	全文検索

るために調査する「無効資料調査」がある。

3.1 検索対象の知識ルール

検索対象の知識ルールを、以下の表 3 に示す。

検索対象妥当性チェック機能は、調査者から入力された調査目的に応じて、表 3 のルールを適用し、検索式の妥当性をチェックする。

3.2 検索対象期間の知識ルール

検索対象期間の知識ルールを、以下の表 4 に示す。

なお、出願前調査において、技術の流れを理解した調査者が、過去 10 年、過去 5 年などに限定するのは許容して良いので、出願前調査のルールにおいて、期間限定が入っていた場合に、上記の旨のメッセージを出力する。

検索対象期間妥当性チェック機能は、調査者から入力された調査目的に応じて、表 4 のルールを適用し、検索式の妥当性をチェックする。

3.3 検索フィールドの知識ルール

検索フィールドの知識ルールを、以下の表 5 に示す。

検索フィールド妥当性チェック機能は、調査者から入力された調査目的に応じて、表 5 のルールを適用し、検索式の妥当性をチェックする。

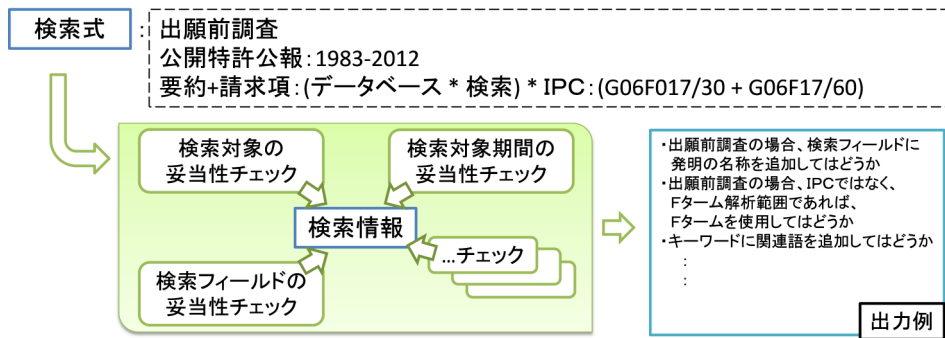


図 6 検索式チェック機能の概念図

Fig. 6 The key map of a search expression check function.

表 6 分類の知識ルール

Table 6 The knowledge rule of classification

調査目的	分類の使用
出願前調査	Fターム / (技術分野により) FI ・「技術分野により」は、実際には「テーマコード」の意味。 ・Fターム解析されていないテーマで、FIを選択する。 ・Fタームを利用する事で、審査官に類似したサーチ戦略を狙う。
実施前調査	FI記号が最優先。 Fタームを漏れ防止用に追加してもよい。
無効資料調査	どんな方法を使っても構わない。

3.4 用語の知識ルール

用語は、調査目的に依存しない。そして、以下のような用語の使い方は適切ではない。

第一に、分類を代表する用語を、その分類と AND 演算している場合である。つまり、データベース技術を示す IPC(G06F 17/39) と用語「データベース」を AND 演算している検索式は妥当ではない。用語「データベース」により、絞り込みの効果はなく、かつ漏れを生じさせる。用語「データベース」と同義語「DB」「データベース」のみを用いている特許書類が検索式にかからないからである。

第二に、その分類に対して、極めて一般的な用語を用いることは適切ではない。例えば、情報技術分野 (例えば、IPC「G06F」) において、用語「情報」「データ」等は、絞り込みの意味を有さず、検索式で用いるべきではない。

第三に、ある用語の同義語等の関連語は、その用語と OR 演算を行うべきである。

上記の第一、第二の知識ルールは、特許コードと、用いることが適切ではない 1 以上の用語との対応表により記述できる。また、第三の知識ルールは、上述した PSA の関連語辞書で実現され得る。

3.5 分類の知識ルール

分類の知識ルールを、以下の表 6 に示す。

上記の表 6 のルールに加えて、2 以上の分類間で AND 演算した場合、検索結果が 0 件となる等、好ましくない場合がある。さらに、共起確率の高い分類は OR 演算することは漏れを少なくするために有効である。

そのため、AND 演算することが妥当でない 2 以上の分類

の集合、共起確率の高い 2 以上の分類の集合をデータベース化する。そして、これらの分類の集合を用いて、分類の妥当性をチェックする。分類の妥当性をチェック機能の概念図を図 7 に示す。

4. 動的な関連用語提案機能

現バージョンの PSA において、関連語辞書を用いて、入力された検索式に含まれる用語の関連語を提案し、用語の漏れを防止する機能を実現している。この機能を静的な関連用語提案機能という。

しかし、PSA の関連語辞書は、一定の記載パターンに合致する関連語を、約 20 年分の特許書類から自動抽出したものである。つまり、一定の記載パターンに合致しない用語は抽出できていない。

そこで、静的な関連用語提案機能を補完するための機能が動的な関連用語提案機能である。動的な関連用語提案機能は、調査者から検索式が入力された後、検索式に含まれず、かつ有用である可能性のある用語を提案する機能である。動的な関連用語提案機能において、まず、検索式から分類の情報を取得する。次に、分類の情報のみを用いて、特許書類を検索する。次に、特許書類の中から、出現頻度が閾値以上の用語を関連語の候補として抽出し、出力する。なお、検索式に検索フィールドが指定されている場合、関連語の候補は、その検索フィールドから抽出する。

例えば、図 8 に示すように検索式「要約+請求項:(データベース * 検索) * FI:(G06F017/30110F + G06F17/30170Z)」が入力された場合、分類「FI:(G06F017/30110F + G06F17/30170Z)」を用いて特許書類を検索し、検索した特許書類の検索フィールド「要約+請求項」から各用語の出現頻度を取得する。そして、出現頻度が N 以上の用語を抽出して提案する。

5. 選別結果を利用した関連特許の提案機能

5.1 理想的検索式を用いた関連特許の抽出

特許調査を実施する場合、上述したように検索式策定工程と選別工程とを行う。選別工程とは、検索式を実行して

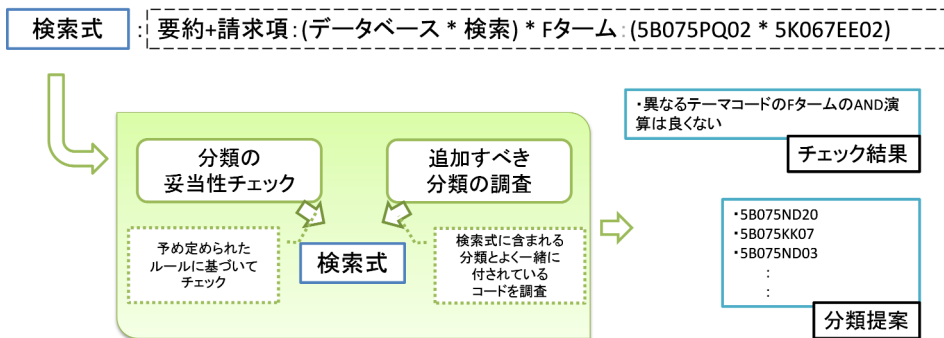


図 7 分類のチェック機能の概念図

Fig. 7 The key map of a classification check function.

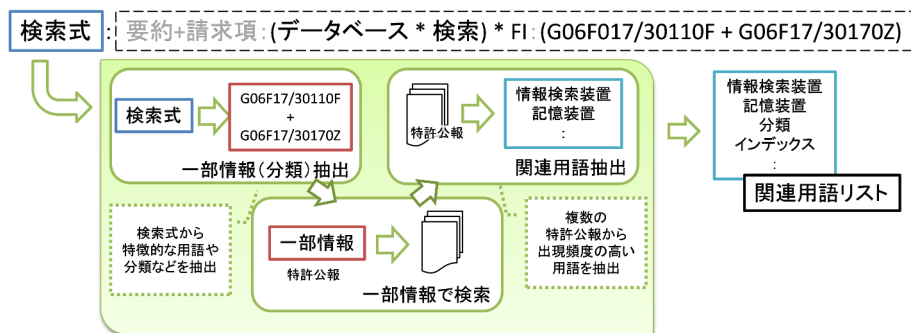


図 8 動的な関連用語提案機能

Fig. 8 The dynamic suggestion function of a related term.

抽出された特許書類を人手で選別し、関連特許と非関連特許とに分ける作業である。

そして、関連特許と非関連特許とに分けた後、関連特許に付された分類と関連特許に含まれる用語とを用いて、すべての関連特許が抽出でき、非関連特許が最大限抽出されない理想的検索式を生成する。理想的検索式の生成のために、ユーザから用いる 1 以上の要素の指定を受け付ける。そして、すべての関連特許が包含される 1 以上の要素の論理和の集合を 1 以上取得する。次に、1 以上の各論理和の集合を実行した場合に、含まれる非関連特許の数が最少の論理和の集合を理想的検索式として決定する。

例えば、ユーザが、「FI」を指定した場合、すべての関連特許が包含される FI の論理和の集合を 1 以上、取得する。そして、1 以上の各論理和の集合を実行し、各論理和の集合ごとに非関連特許の数を算出する。次に、非関連特許の数が最少の論理和の集合を理想的検索式として決定する。

理想的検索式が得られれば、その次に、「理想的検索式—入力した検索式」を実行し、漏れている可能性のある特許書類を抽出する。

5.2 選別した結果を用いた関連特許の抽出

まず、選別した関連特許の特許書類から多数の自立語(用語)を抽出し、特許書類ごとにベクトルを生成する。また、非関連特許の特許書類から多数の自立語を抽出し、特

許書類ごとにベクトルを生成する。その際、関連語辞書を用いて、関連語は同一の用語として集約し、ベクトルを生成する。

次に、非関連特許の中で、関連特許との距離が近く、他の非関連特許との距離が遠い非関連特許を、関連特許の候補として提示する。

また、関連特許の中で、他の関連特許との距離が遠く、非関連特許との距離が近い関連特許を、非関連特許の候補として提示する。

なお、特許書類から用語を抽出する場合、抽出するフィールドを実験により検証する必要があると考える。

6. 関連業績

HYPAT-i[4], PATOLIS[5], NRI サイバーパテントデスク 2[6] などの特許調査ツールの多くは、概念検索という機能を有する。概念検索は、検索対象の技術を説明する文を入力すると、入力した文が示す技術的概念に関連の深い特許公報を検索する機能であり、特許調査の非専門家が利用しやすい検索機能である。しかし、概念検索は、システム内部の処理が外からは見えず、調査方法の良し悪しの判断、調査結果の評価、および調査結果の再利用が困難である。そこで、一般的に、サーチャーは、調査対象に関連したキーワードと、その関連語や分類で構成される検索式を用いる。しかし、上述したように、検索式の作成は、調査

者の経験とスキルに大きく依存し、サーチャーでも容易ではない。

7. おわりに

キーワードを入力するだけで関連語と分類を提案する PSA により、大幅な特許調査フェーズの改善が見られた。

今後、今回提案の機能である、サーチャーの知識を利用した機能を実装し、評価を行っていく。

参考文献

- [1] 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, pp. 458-465 (2008).
- [2] 谷川英和, 日本知財学会 第 8 回年次学術研究会発表会, 特許調査の検索式提案システム.
- [3] 特許電子図書館 (オンライン), <http://www.ipdl.inpit.go.jp/homepg.ipdl>.
- [4] 特許検索システム HYPAT-i (オンライン), <https://app01.hypatweb.jp/hypati/>.
- [5] 特許検索システム PATOLIS (オンライン), <http://www.patolis.co.jp/>.
- [6] 特許検索システム NRI サイバーパテントデスク 2 (オンライン), <http://www.patent.ne.jp/>.