

特許検索式提案システムについて

－ CPC に対応した検索式の提案－

Introduction of Query Proposal System for Patent Search

IRD 国際特許事務所 所長／弁理士／有限会社アイ・アール・ディー **谷川 英和**

PROFILE 1986年神戸大学工学部システム工学科卒業。同年、松下電器産業（株）に入社し、中央研究所等において、データベース管理システム等の研究開発に従事。1999年弁理士試験合格。2002年1月、IRD 国際特許事務所を開設。所長、弁理士。2003～2007年3月京都大学 COE 研究員、2007年4月～京都大学非常勤講師、2009年4月～東京工業大学客員教授。博士（情報学）。弁理士会、日本知財学会、情報処理学会各会員。2007年度から特許版・産業日本語委員会委員。

✉ htanigawa@ird-pat.com ☎ 06-6944-4530

豊橋技術科学大学工学部情報・知能工学系研究員 **太田 貴久**

PROFILE 2006年豊橋技術科学大学大学院博士前期課程知識情報工学専攻修了。現在、同大学研究員。修士（工学）、言語処理学会会員、日本知財学会会員。2014年度から特許版・産業日本語委員会委員。

✉ t-ota@t-ota.com

1 はじめに

特許調査を行う者（以下、調査者とする）は、膨大な数の特許公報から調査対象となる特許公報を検索する。検索時には、調査対象に関連したキーワードだけではなく、調査対象を適切に取得するために、キーワードと、その同義語、上位語、下位語（以下、これらを類義語とする）や、IPC、F ターム（以下、これらをコードとする）などの分類コードで構成される検索式を用いることが一般的である。しかしながら、検索式の作成は、調査者の経験とスキルに大きく依存し、かつ、適切な検索式でなければ有効な検索ができないという問題がある。

特許公報は、同一の概念の用語であっても、明細書の作成者が違えば異なる表現が用いられている。例えば、「シミュレータ」と「疑似体験装置」のような場合である。特許検索において同じ概念の用語が記載された特許公報を漏れなく収集するためには、表現の揺れを考慮した類義語を用いる必要がある。しかし、特許公報は、「疑似体験装置」のような一般的な類義語辞書に登録されていないような用語が多用されているため、経験の浅い調査者であれば適切な類義語を用いることができない。

また、特許公報は、それぞれに分野を示すコードが割

り当てられている。特許検索において調査対象の分野の特許公報のみに限定するためには、適切なコードを用いる必要である [1]。しかしながら、IPC は、約 7 万種類、F タームは、約 36 万種類に分類されているため、経験の浅い調査者であれば検索結果を適切に絞り込むことができない。さらに、国際的な特許検索をする場合、調査者は、各国で導入が進められている新しい分類コードである CPC を考慮する必要がある。

以上より、特許検索に特化した類義語辞書を構築し、類義語辞書を用いて、キーワードとその類義語、コードで構成された検索式を提案する検索式提案システムを開発した。なお、本システムは、簡易的ながら、国内公報についても CPC を含む検索式を提案できる機能を有する。

2 検索式提案システムの概要

本システムの概要を図 1 に示す。図 1 のように、本システムは、大きく分けると類義語辞書構築部と検索式構築部の 2 つで構成される。

類義語辞書構築部は、特許公報を解析して、類義語を収集する。また、検索式構築部は、ユーザが入力したキー

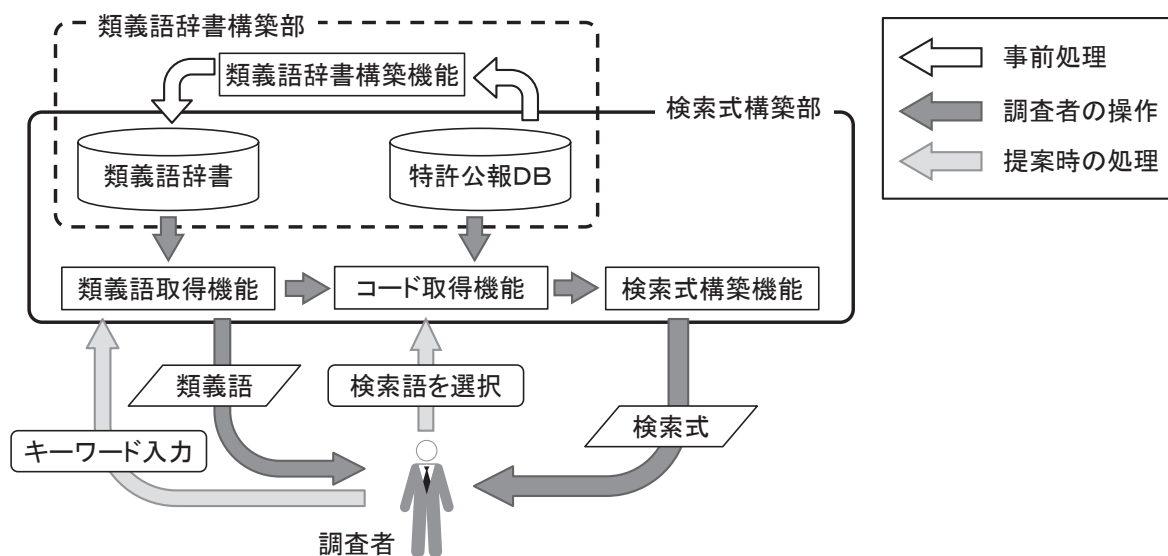


図1 システムの概要

ワードの類義語を提示したり、各類義語に関連する特許分類コードを提示することで、ユーザの検索式構築を補助する。以下、各部の概要について説明する。

2.1 類義語辞書構築部

類義語辞書構築部では、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、類義語辞書を構築する。類義語辞書構築部では、難波らの研究 [2] をもとに、公報から類義語を抽出する。

難波らは、「などの」や「等の」という表現の前後に記載されている用語は上位語、下位語関係にあることに着目した。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」の上位語の関係にある。しかしながら、「などの」と「等の」だけでは、「パソコンなどのキーボード」のように上位語、下位語関係にない場合であっても、誤って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、上位語、下位語関係にないとした。また、「メモリやHDDなどの記録媒体」のように共通の上位語、もしくは下位語を持つ用語が列記されている場合も上位語、下位語関係に

ないとした。さらに、共通の上位語と下位語を持つ用語は、同義語であるとした。

類義語辞書構築部は、以上の難波らの手法を用いて実装した。以下、その手順を示す。

- (i) 特許公報から「A などの B」と「A 等の B」という表現を抽出し、用語の対を取得する。
- (ii) 特許公報から「A の B」という表現を抽出し、用語の対を取得する。
- (iii) 特許公報から「A や B などの～」という表現を抽出し、用語の対を取得する。
- (iv) i で抽出した用語の対から、ii と iii で抽出した用語の対を削除する。
- (v) iv で削除した「A」を下位語、「B」を上位語として類義語辞書に登録する。
- (vi) v で構築した類義語から同一の上位語と同一の下位語を持つ用語の対を取得する。
- (vii) vi で取得した用語の対を同義語として類義語辞書に登録する。

2.2 検索式構築部

検索式構築部は、調査者の入力したキーワードから類



義語を取得する類義語取得機能と、キーワードとその類義語（以下、これらを検索語とする）から適切な分類コードを取得するコード取得機能、検索語とコードで構成された検索式を作成する検索式構築機能で構成される。

類義語取得機能

本機能は、キーワードに対応した類義語を類義語辞書から取得する。以下、その手順を示す。

- (i) 類義語辞書からキーワードの同義語を類義語辞書から取得する。
- (ii) キーワードと i で取得した同義語の上位語、下位語を類義語辞書から取得する。

コード取得機能

本機能は、キーワードと、類義語取得機能が取得した類義語から調査者が調査の目的に応じて選別した検索語を取得し、それらの検索語と関連の強い分類コードを取得する機能である。本機能では、特許公報ごとに検索語の出現率（公報内の全用語のうち、検索語の占める割合）を算出し、コードごとにそのコードを保持する特許公報の用語の出現率の和（以下、コード重要度とする）を求め、コード重要度が高いほど適切なコードであるとした。以下、その手順を示す。

- (i) 検索語を取得する。
- (ii) i で取得した検索語からキーワードとその類義語からなるグループを 1 以上作成する。
- (iii) ii で作成したグループ内の用語を「OR」で連結する。
- (iv) iii で連結したグループ間を「AND」で連結し、検索式を作成する。
- (v) iv で作成した検索式を用いて特許公報 DB から特許公報を取得する。
- (vi) v で取得した特許公報の「検索語の出現回数 / 特許公報内の全用語数」から、用語の出現率を算出する。
- (vii) iv で取得した特許公報から IPC と F タームを取得し、コードごとにそのコードを保持する特許公報の用語の出現確率の和からコード重要度を算出する。
- (viii) vii で算出したコード重要度の総和を算出する。
- (ix) vii で算出したコード重要度を値が大きい順に加算し、「加算した値 / コード重要度の合計」が閾値を越えたとき、加算したコード重要度に対応したコードを取得する。

検索式構築機能

本機能は、検索語とコード取得機能で取得したコードを用いて、検索語とコードで構成された検索式を調査者へ提案する。以下、その手順を示す。

- (i) コード取得機能 i から iv の手順で作成した検索式を取得する。
- (ii) i で取得したコードが複数である時、全てのコードの組み合わせを作成する。
- (iii) i で取得した検索式に対して、ii で作成した組み合わせを「AND」で連結し、組み合わせの数だけ検索式を構築する。

コード取得部における CPC への対応

最後に、はじめにでも述べたように、本システムは CPC に対応している。具体的には、コード取得部において、IPC や F タームに加え、CPC も関連コードとして出力する機能を有する。以下、その手順を示す。

- (i) コード取得部の ix で取得した IPC コードを集める。
 - (ii) 特許庁が公開している分類対照ツール [3] を元に、IPC コードに対応する CPC コードをすべて取得する。
- 上記のとおり、本システムの CPC は、単純な対応表に基づくものである。そのため、1 つの IPC に対して関連の弱い多くの CPC が対応付けられてしまう問題がある。例えば、IPC 「G06F 17/27」は自然言語処理（我々が日常で使う言語をコンピュータで処理する分野）に関するコードだが、ここには、言葉の構造を解析する分野（CPC 「G06F 17/2705」等）や、辞書に関する分野（CPC 「G06F 17/2795」）といった多くの処理が含まれる。このため、現在、「辞書」という検索語に対して、CPC 「G06F 17/2705」等も出力してしまう。

この問題については、今後、F タームや明細書本文等を用いて、より検索語に適した CPC を取得できるような改善を予定している。

3 検証

本システムの有用性について評価するため、提案する類義語の有用性と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、特許情報プラットフォーム

フォーム（J-PlatPat）で行い、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、適合公報とする）であるかの確認は、検索結果から人が目視で確認したものである。

3.1 提案する類義語の有用性

本システムが提案する類義語が、同一の概念でかつ、異なる表現で記載された特許公報を検索する場合に有用であるかについて、検索語で構成された検索式を用いて検索した結果に含まれる適合公報の件数（以下、適合数とする）が向上することを確認する。

本検証では、「OS」、「タスク」、「速度」をキーワードとする調査を例に挙げる。キーワードで構成された検索式で検索した結果を表1に示す。

本システムの類義語取得機能を用いて、キーワードに対する類義語を取得した結果を表2に示す。

本検証では、表2の上位語、下位語が調査対象を必要以上に拡大してしまう判断したため、同義語のみを選

択し、検索式を作成した。検索語で構成された検索結果を表3に示す。

表1と表3より、本システムの提案する類義語を用いることで適合数が14件から54件となり適合公報を多く取得できたことが分かる。

3.2 提案する検索式の有用性

本システム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4に示す。なお、本検証では、コードとしてIPCを用いた。

表4より、本システムが提案する検索式を用いることで、表3の結果に比べ検索結果に対する適合公報の割合を向上させたことがわかる。

表1 キーワードで構成された検索式の検索結果

検索式	検索結果	適合数
(OS) * (タスク) * (速度)	60	14

表2 類義語の取得結果

キーワード	種別	類義語
OS	同義語	OperatingSystem オペレーティングシステム 基本ソフト 基本ソフトウェア
	上位語	システム
	下位語	DOS UNIX Linux リナックス ……
タスク	同義語	プロセス ジョブ スレッド task
	上位語	処理単位 実行単位
速度	同義語	スピード
	下位語	回転速度 伝播速度 落下速度 ライン速度 ……

表3 検索語で構成された検索式の検索結果

検索式	検索結果	適合数
(OS+ オペレーティングシステム +OperatingSystem+ 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)	681	54

表4 検索語とコードで構成された検索式の検索結果

検索式	検索結果	適合数
(OS+ オペレーティングシステム +OperatingSystem+ 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)	45	41

4 おわりに

narabe.cgi (2015年8月31日確認)

本システムを用いることにより、経験の浅い調査者であっても、キーワードから適切な検索語とコードで構成された検索式を作成することができることを実証した。しかしながら、本システムは、類義語を用いることで多くの適合公報が取得できたにもかかわらず、提案した検索式では、検索語から構成された検索式で取得できていた適合公報の取得漏れが発生した。

最後に、本システムに関する今後の予定を示す。

- CPCの精度向上

現在のシステムでは、単純にIPCをCPCに展開したものを利用しているため、無関係の（もしくは、関連の少ない）CPCが提示されてしまう問題がある。そこで、Fタームや明細書の文章を利用して、より検索語に適したCPCを提案できるようにする。

- 漏れのない検索式の提案

本稿の実験より、検索語のみで検索した場合に取得できていた特許が、検索語＋分類コード（IPC）で検索すると漏れてしまう問題がある。そこで、分類コードを条件に追加すると漏れてしまう特許の持つキーワードをユーザに提示する等の機能を実装し、検索漏れの可能性を極力減らす。

- インターフェースの改善

近年、Webブラウザの機能が向上している。これらの機能を活用し、より直感的に検索式を構築できるようにインターフェースを改善する。

参考文献

- [1] 谷川英和, 河本欣士: 特許工学入門—発明の着想から特許権取得までのプロセス論, 中央経済社, pp.76-77, 2003
- [2] 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ソースの自動構築, 言語処理学会, 第13回年大会, pp.1113-1116, 2007
- [3] 特許庁, 分類対照ツール, https://www.jpo.go.jp/cgi/cgi-bin/search-portal/narabe_tool/

