

特許オントロジーに基づく類義語辞書を利用した 特許検索式構築支援システム

谷川英和^{†1} 太田貴久^{†2}

概要：特許書類特有の表現を考慮して、特許書類に出現する用語から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特許検索式構築支援システムを開発した。

キーワード：特許調査、検索式、特許オントロジー、類義語辞書

Patent search expression construction support system by using a synonym dictionary based on patent ontology

HIDEKAZU TANIGAWA^{†1} TAKAHISA OTA^{†2}

Abstract: We constructs a synonym dictionary specialized patent search from terms appearing in the patent documents in view of the patent documents-specific expression, and we have developed a patent search expression construction support system that proposes search expressions configured keywords ,their synonyms, and the codes by using the synonym dictionary.

Keywords: patent search, search expression, patent ontology, synonym dictionary

1. はじめに

我々は、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きている。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。

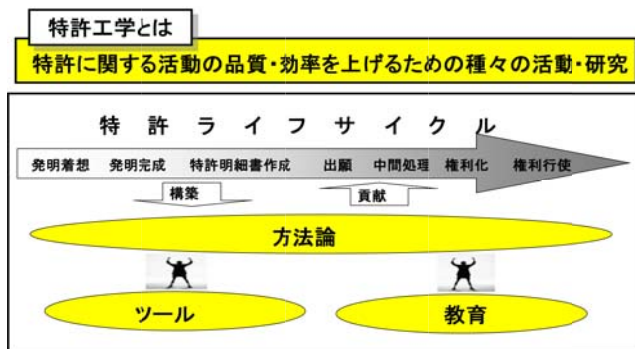


図1：特許工学の領域

Figure 1 Disciplines of patent engineering

一方、従来から、特許調査ツールをはじめ、特許ライフサイクルにおける各種作業を支援するツールは多数、存在

している。特に、特許調査ツールは、HYPAT-i, PATOLIS, NRI サイバーパテント、特許庁の J-PlatPat 等、多数存在する。

また、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書類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が、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、特許調査の専門家は、先行特許調査において、まず、調査対象の技術に対して、1以上の主題を決定し、主題に対応する用語を検討する。次に、各主題を示す同義語、上位語、下位語（以下、これらを類義語という）を決定し、最後に、IPC、F ターム等の特許分類コード（以下、「コード」という。）を選定する。そして、最終的に、図2に示すような検索式を得る。

しかし、検索式の作成は、調査者の経験とスキルに大きく依存し、特許書類に記載されている適切なキーワードの選択、および適切なコードの選択は、特許調査の専門家でも容易ではないという問題がある。

つまり、特許請求の範囲、特許明細書を有する特許書類（公開特許公報、特許公報）は、同一の概念の用語であっても、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書

^{†1} IIRD 国際特許事務所・(有)アイ・アール・ディー
IRD Patent Office, IRD corp.
^{†2} 豊橋技術科学大学
Toyohashi University of Technology

類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、特許書類は、「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。

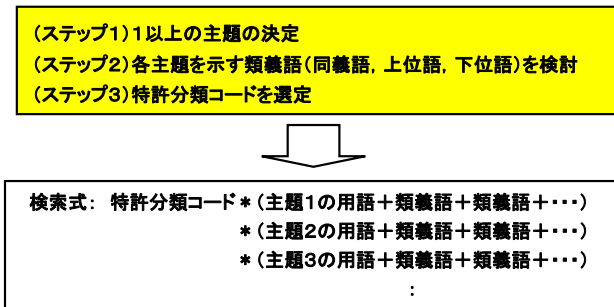


図 2：検索式作成の標準的な手順

Figure 2 Standard procedures of creating search expression

また、特許書類は、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書類のみに限定するためには、適切なコードを用いる必要がある。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。

以上のような背景のもと、特許検索に特化した類義語辞書を構築し、当該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特許検索式構築支援システム (PatentSearchAssistant*3)) を開発した。なお、本システムにおいて、国内公報についてもCPC (Cooperative Patent Classification) を含む検索式を提案できる機能をも有する。なお、CPCは、EPO (欧州特許庁) とUSPTO (米国特許商標庁) との間で合意された特許分類であり、両庁の分類付与業務を軽減することや、調査業務を効率化することなどを目的としおり、日本においても導入される予定である4)。

2. 特許検索式構築支援システムの概要

本システムの概要を図3示す。本システムは、類義語辞書構築部と検索式構築部の2つより構成される。

類義語辞書構築部は、特許書類を解析して、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により、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。

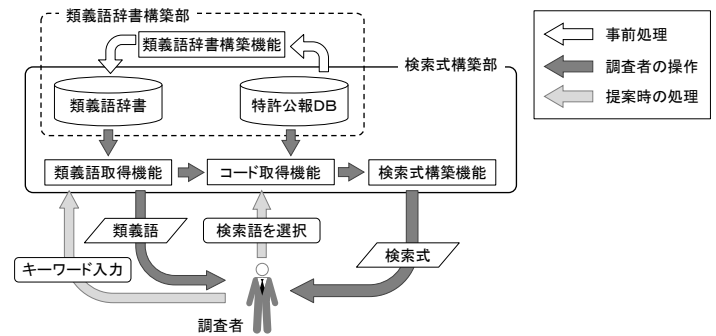


図 3 特許検索式構築支援システムの概要

Figure 3 Overview of patent search expression construction support system

(1) 類義語辞書構築部

類義語辞書構築部では、特許書類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、難波らの研究成果5)の利用および特許請求の範囲の解析処理により、特許書類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。

また、特許請求の範囲を構成する請求項のうち、内的付加の請求項の表現 (例えば、「前記記録媒体はHDDである請求項1記載の情報処理装置。’)から、「記録媒体」が「HDD」の上位語の関係にある、と判断し、上位語および下位語の登録数を増加させた。

類義語辞書構築部は、以上の手法を用いて実装した。

(2) 検索式構築部

検索式構築部は、主に、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、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、および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からなる。また、付加的な機能として、検索式構築部は、コードの選択支援機能と保存機能を有する。コードの選択支援機能は、コードの意味を説明するコード説明出力機能、構築した検索式により検索される代表的な特許書類を出力する参考公

報出力機能を有する。また、保存機能は、ユーザが追加した類義語や、削除した類義語、選択した類義語などを保存する機能である。保存機能は、追加・削除類義語保存機能、選択類義語保存機能、選択コード保存機能、検索式保存機能から構成される。

(2-1) 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。本機能では、2階層までの類義語を出力できる。

以下の図4は、ユーザが入力したキーワードに対する類義語の出力例である。図4において、左側はキーワード「検索」に対する類義語の出力例、右側がキーワード「検索」の類義語「探索」に対するさらなる類義語であり、「リサーチ」等の用語が出力されている。



図4 類義語の出力例

Figure 4 Output example of synonyms

(2-2) コード取得機能

本機能は、ユーザが入力した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用語の集合である検索語を取得し、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、特許書類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書類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書類DBから特許書類を取得する。
- (vi) vで取得した特許書類の「検索語の出現回数/特許書類内の全用語数」から、用語の出現率を算出する。
- (vii) ivで取得した特許書類からIPCとFターム等のコードを取得し、コードごとにそのコードを保持する特許書類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。

(2-3) 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（図5参照）。



図5 検索式の例

Figure 5 example of search expression

(2-4) 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6)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」）といった多くの小分野が含まれる。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。

なお、以下の図 6、図 7 がコードの出力例である。

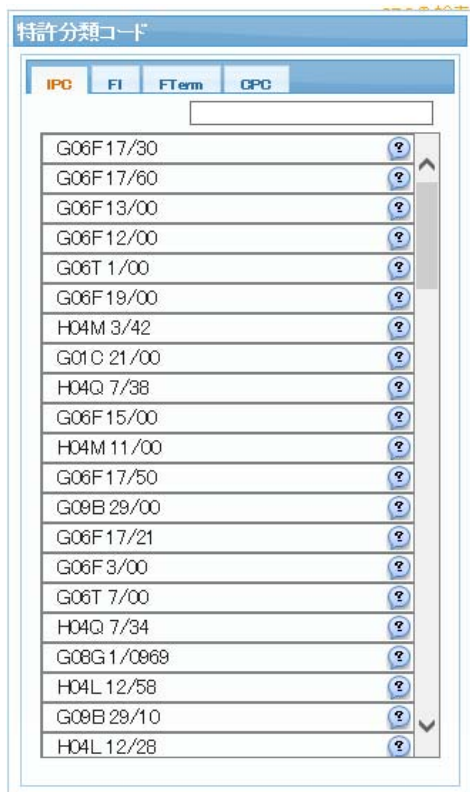


図 6 コードの出力例 (1) IPC

Figure 6 Output example of code (1) IPC

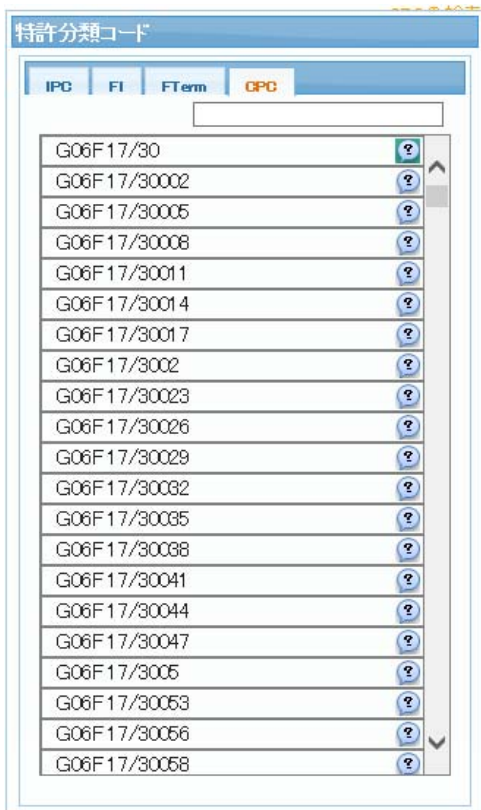


図 7 コードの出力例 (2) CPC

Figure 7 Output example of code (1) CPC

(2-5) 特許分類コードの選択支援機能

ユーザが選択したコードが正しいものであるか否かを判断するために、検索式構築部は、抽出したコードの意味を出力するコード説明出力機能、および選択した特許分類コードが付されている特許公報を提示する参考公報出力機能を有する。コード説明の出力例を図 8 に示す。

また、コードの説明は、ユーザにとって理解し難い場合もある。そこで、検索語と選択したコード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参考公報出力機能も開発した。参考公報のリストの表示例を図 9 に示す。また、ユーザが選択した公報の表示例を図 10 に示す。



図 8 コード説明の出力例

Figure 8 Output example of descriptions of codes

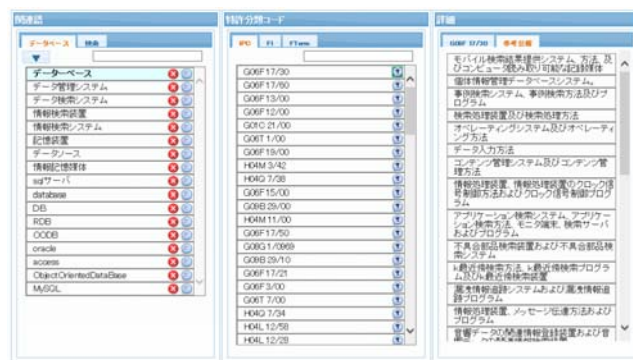


図 9 参考公報のリストの表示例

Figure 9 Display example of the list of patent publications

提示される特許公報のリストでは、同じようなコード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、コードが 2 つ付された特許公報とコードが 3 つ付された公報があった場合、付されているコード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、ユーザは、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、コードを選択することができる。

そして、図6において、ユーザは、表示されている類義語のうち、任意の類義語を指定し、提案されたコードのうち、「G06F 17/30」を選択したとする。すると、当該類義語と、コード「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。公報の本文の表示例を図7に示す。

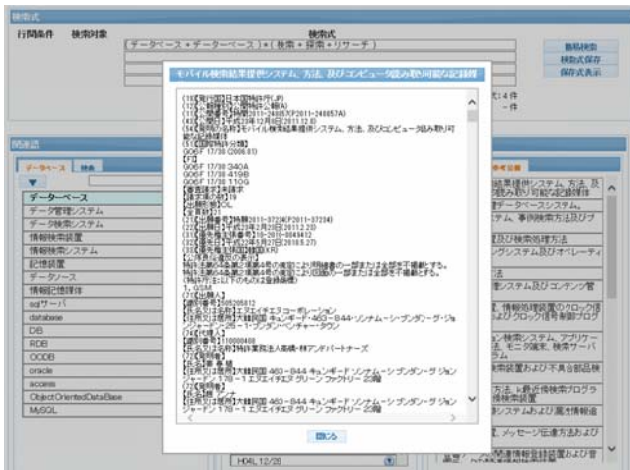


図10 参考公報の表示例

Figure 10 Display example of a patent publication (2-6)保存機能

保存機能を構成する追加・削除類義語保存機能は、ユーザが任意に追加・削除した類義語を保存する機能である。追加・削除類義語保存機能の使用例を図11に示す。

図11において、ユーザは、「情報格納部」を追加し、「情報検索装置」を削除している。追加した「情報格納部」は太字で表示され、削除した「情報検索装置」は表示されなくなる。

また、選択類義語保存機能は、ユーザが選択した類義語（最終的に、検索式に使用した類義語）を保存する機能である。選択類義語保存機能の使用例を図12に示す。

図12において、ユーザは、入力したキーワード「データベース」に対して提案される類義語のうち、「データ管理システム」と「情報検索装置」を選択し、検索式を構築・保存したとする。再度、キーワード「データベース」を入力すると、以前に選択した「データ管理システム」と「情報検索装置」は、下線付きで表示され、かつ上位に表示される。



図11 追加・削除類義語保存機能

Figure 11 Save function of addition and deletion synonyms



図12 選択類義語保存機能

Figure 12 Save function of the selection synonyms

さらに、検索式保存機能は、ユーザが構築した検索式を保存する機能である。検索式保存機能の使用例を図13に示す。

図13において、ユーザは、「要約+請求項」に「(データベース + データ管理システム) * (検索)」, 「IPC」に「G06F 17/30 + G06F 17/60 + G06F 12/00」を指定した検索式を保存している。再度、保存した検索式よりも検索範囲が広くなるような検索語、もしくはコードを選択すると、以前に保存した検索式が表示される。



図13 検索式保存機能

Figure 13 Save function of the search expression

3. 検証

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。また、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。さらに、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

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

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

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。

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, 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) 間瀬久雄: 特許を対象とした概念検索の技術課題, Japio YEAR BOOK 2010 200-207, 2010
- 3) 谷川英和, 渡辺俊規, 増満光, 難波英嗣: 特許調査の検索式提案システム, 日本知財学会第8回年次学術研究発表会, 2G6, 2010
- 4) 酒井美里: CPC(共通特許分類)の始動と特許情報検索への影響, Japio YEAR BOOK 2013 208-213, 2013
- 5) 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ラスの自動構築, 言語処理学会, 第13回年大会, pp.1113-1116, 2007
- 6) 特許庁, 分類対照ツール, https://www.jpo.go.jp/cgi/cgi-bin/search-portal/narabe_tool/narabe.cgi (2015年8月31日)