

# 特許調査の高品質化・高効率化のための特許検索式の自動提案技術

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年～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、2011年4月～大阪大学非常勤講師 博士（情報学）。弁理士会、情報処理学会各会員。

✉ htanigawa@ird-pat.com

TEL 06-6944-4530

IRD 国際特許事務所／有限会社アイ・アール・ディー **渡辺 俊規**

## PROFILE

2004年奈良工業高等専門学校専攻科電子情報工学専攻修了。2006年京都大学大学院情報学研究科修了。2004年ベンチャー企業設立を経て、2008年IRD 国際特許事務所入所。修士（情報学）。日本知財学会会員。

IRD 国際特許事務所／有限会社アイ・アール・ディー **増満 光**

## PROFILE

2008年関西大学大学院入学。2008年8月IPA：2008年度上期未踏IT人材発掘・育成事業（未踏コース）開始、特許請求項の参照関係可視化に関する研究開発に従事、2009年3月IPA：2008年度上期未踏IT人材発掘・育成事業（未踏コース）終了、2010年関西大学大学院卒業、同年2010年IRD 国際特許事務所入所。修士（情報学）。

広島市立大学大学院情報科学研究科准教授 **難波 英嗣**

## PROFILE

1996年東京理科大学理工学部電気工学科卒業。2001年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。同年、日本学術振興会特別研究員。2002年東京工業大学精密工学研究所助手。同年、広島市立大学情報科学部講師。2010年4月広島市立大学大学院情報科学研究科准教授、現在に至る。博士（情報科学）。言語処理学会、情報処理学会、人工知能学会、ACL、ACM 各会員。

## 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た（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作業において、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、特許調査の専門家は、適切

な特許国際分類（IPC）、FI、またはFターム（以下、これらを総称して「特許分類コード」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、特許調査の専門家でも、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、特許調査の非専門家にとって、特に、適切に特許分類コードを決定することは困難である。

そこで、我々は、ユーザが入力した日本語のキーワードに対応する関連語と特許分類コード（IPC、FI、およびFターム）を提案し、選択された関連語と特許分類コードから検索式を構築する特許調査支援システム（PatentSearchAssistant、以下、PSA）を開発した。

PSAにより、ユーザは、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。

また、PSAを米国特許に拡張した米国版PSAの試作版も開発したので、報告する。

## 2 関連業績

HYPAT-i、PATOLIS、NRIサイバーパテント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、特許調査の専門家は、調査対象に関連したキーワードと、その関連語や特許分類コード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、特許調査の専門家でも容易ではない。

## 3 特許調査支援システム (PSA) の概要

PSAは、キーワードを入力するだけで、特許公報に記載されている関連語とIPCなどの特許分類コードを用いた検索式を提案する。この検索式は、調査専門家が作成する検索式と同様に品質が高い。

本システムは、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る。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、および特許分類コードの提案の画面例を、図2に示す。

図2は、ユーザ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は、ユーザが入力したキーワード、およびユーザが選択した関連語（以下、これらを検索語とする）との関連性が高い特許分類コードを提案している画面である。

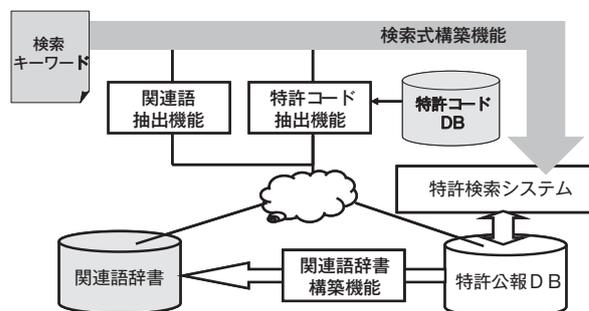


図1 PatentSearchAssistantの概要



図2 関連語・特許分類コードの提案

## 4 PSA の各種機能

### 4.1 関連語辞書構築機能

関連語辞書構築機能は、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<sup>2)</sup>。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。

### 4.2 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、適切な特許分類コードを抽出する特許コード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、ユーザはキーワード「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ユーザが選択した関連語である。

次に、特許コード抽出機能の出力例を図4に示す。



図3 関連語抽出機能の出力例

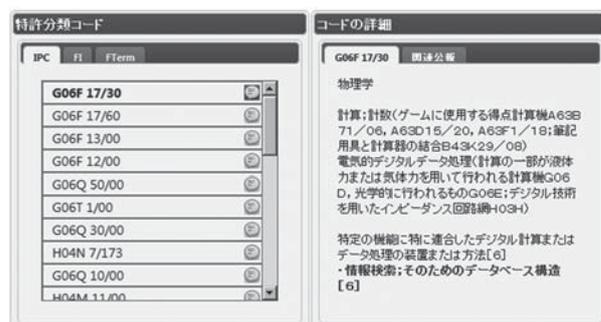


図4：特許コード抽出機能の出力例

図4において、特許コード抽出機能は、関連語抽出機能で、ユーザが選択した検索語を用いて、検索に適した特許分類コードを抽出する。抽出する特許分類コードは、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特許分類コードにその割合を重みとして与え、特許分類コードごとに重みの総和（以下、コード重要度という）を算出し、コード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、取得した特許分類コードの組み合わせと検索語（ユーザ入力のキーワードと、ユーザ選択の関連語）から構成される。この時、ユーザは使用する特許分類コードを任意に選択し、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。

### 4.3 特許分類コードの選択支援機能

ユーザが選択した特許分類コードが正しいものであるかを判断するために、PSAは、抽出した特許分類

コードの意味を出力する特許分類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関連公報提示機能を有する。

特許分類コード説明の出力例を図4に示す。特許分類コード説明出力機能では、各特許分類コードの説明を出力する。この説明は、ユーザにとって理解し難い場合もある。そこで、検索語と選択した特許分類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 関連公報提示機能

図5において、ユーザは、任意の検索語を指定し、提案された特許分類コードのうち、「G06F 17/30」を選択したとする。すると、当該検索語と、特許分類コード「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6に示す。



図6 公報の本文の表示例

この時提示される特許公報は、同じような特許分類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、特許分類コードが2つ付された特許公報と特許分類コードが3つ付された公報があった場合、付されている特許分類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、特許分類コードを選択することができる。

#### 4.4 保存機能

保存機能は、ユーザが追加した関連語や、削除した関連語、選択した関連語などを保存する機能である。保存機能は、追加・削除関連語保存機能、選択関連語保存機能、選択コード保存機能、検索式保存機能から構成される。

追加・削除関連語保存機能は、ユーザが任意に追加・削除した関連語を保存する機能である。追加・削除関連語保存機能の使用例を図7に示す。

図7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。

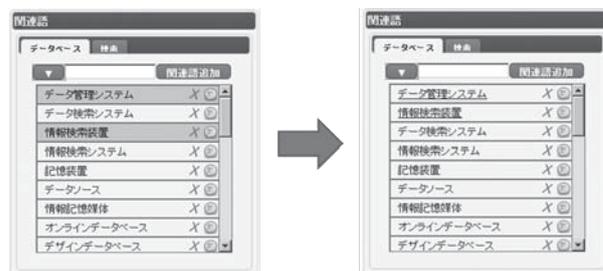


図7 追加・削除関連語保存機能

また、選択関連語保存機能は、ユーザが選択した関連語（最終的に、検索式に使用した関連語）を保存する機能である。選択関連語保存機能の使用例を図8に示す。

図8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関連語のうち、「データ

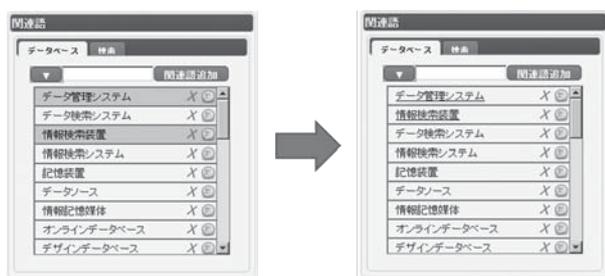


図8 選択関連語保存機能

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。

また、選択コード保存機能は、ユーザが選択したコード（最終的に、検索式に使用したコード）を保存する機能である。選択コード保存機能の使用例を図9に示す。

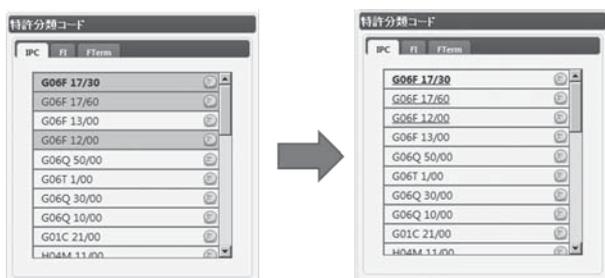


図9 選択関連語保存機能

図9において、ユーザは、入力および選択した検索語「データベース」、「検索」、「データ管理システム」に対して提案されるコードのうち、「G06F 17/30」と「G06F 17/60」と「G06F 12/00」を選択し、検索式を構築・保存したとする。再度、検索語「データベース」、「検索」、「データ管理システム」を入力および選択すると、以前に選択した「G06F 17/30」と「G06F 17/60」と「G06F 12/00」は、下線付きで表示され、かつ上位に表示される。

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10に示す。



図10 検索式保存機能

図10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」を指定して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がるような検索語、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。

#### 4.5 共有機能

共有機能は、追加した関連語や、選択した関連語など、保存機能で保存した情報を、同じ企業や部署などの同じグループ内のユーザ間で共有する機能である。共有機能の使用例を図11に示す。

図11において、ユーザは、キーワード「データベース」と「検索」を入力している。この時、同じグループ内の他のユーザが、キーワード「データベース」と「検索」に対する関連語を追加、削除していたり、これらのキーワードを含む検索式を保存していたりすれば、各保存機能と同様の条件で、関連語やコード、また検索式が提案される。他のユーザが保存した情報の提案の際には、各要素の先頭に「\*」が付加される。



図11 共有機能

#### 4.6 PSAの検証

「OS」、「タスク」、「速度」をキーワードとする調査を例に、特許電子図書館 (IPDL) を用いて、本ツール

の検証を行った。

検索式「要約+請求の範囲：OS \* タスク \* 速度」による検索結果は 41 件であり、そのうち適合特許は 13 件であった。適合特許とは、調査対象として適切な公報であり、人手で特許公開公報を読んで判断した。

一方、本ツールが生成した検索式「要約+請求の範囲：(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)」による検索結果は 39 件であり、そのうち適合特許は 35 件であった。本ツールにより、調査対象数が 41 件から 39 件に減少し、適合数が 13 件から 35 件に増加しており、極めて有効な検索式が提案できたことが分かる。

## 5 米国版 PSA の試作

### 5.1 英語版関連語辞書構築機能

英語版関連語辞書には、二つの手法を用いて関連語を登録した。一つは、日英機械翻訳機を用いて、日本語の関連語辞書に登録されている全ての用語を英語に翻訳し、取得した関連語のうち、実際に米国特許公報で使用されている用語を登録した。もう一つは、日本語版関連語辞書を構築した時と同様に米国特許公報から「such as」などの手がかりになる句を用いて、その前後にある名詞句を取得し、取得した名詞句の同義語、上位概念語、下位概念語などの関係性を考慮して登録した。

### 5.2 米国版検索式構築のための機能

米国版 PSA には、FI、F タームの代わりに US クラスを提案するようにした。提案する特許分類コードの種類以外は、日本版と同様の機能を実装した。米国版 PSA の検索式の構築例を図 12 に示す。図 12 において、ユーザは、キーワード「Database」と「Search」を入力し、各抽出機能が出力した関連語、コードを、その説明文を見ながら選択し、米国用の検索式を構築している。

本システムを利用することで、米国特許に対する調査が不慣れであっても、簡単に検索式を構築することができる。



図 12 米国版 PSA の検索式の構築例

## 6 おわりに

以上、非専門家でも、専門家並みの特許調査を可能にする特許調査支援システム (PSA) および PSA の効果について説明した。また、PSA の保存機能や共有機能を用いれば、特許調査における個の知の蓄積と活用、および、組織的な知の蓄積と活用が可能になると考える。これにより、より効率良く、また、より質の良い特許調査を行うことができる。

さらに、試作した米国版 PSA についても説明した。今後、米国版 PSA の精度を向上させるとともに、ニーズがあると考えられる中国語版 PSA の開発も行いたい。

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。

- 1) 谷川英和他, 特許工学入門, p1 ~ p7(2003), 中央経済社
- 2) 難波英嗣他, 「特許、論文データベースを統合した検索環境および動向分析ツールの構築」, 2006 年度 JAPIO 「特許情報活用の時代の検索と機械翻訳技術」, pp116-119, 2006.