

# 特許調査から特許明細書作成を 統合的に支援する特許構築システム

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

**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 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

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。

✉

TEL

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）。

✉

TEL

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。

✉

TEL

## 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。

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。

しかし、特許ライフサイクルにおける各作業は有機的に繋がっているにも関わらず、既存ツールは、通常、単一のフェーズを支援する単体ツールであり、多数のツールが連携して、連続した複数作業をサポートする統合システムが存在しない。

そこで、本論文において、約15年分（約500万件）の特許公報（特許公開公報、特許公報、実用新案公開公報、実用新案登録公報など）から、弁理士等の専門家の知識、特許法、特許審査基準等の法律事項を考慮した自然言語処理により抽出した関連語辞書、および特許部品

DB を用いて、複数のツールが連携して、統合的に特許ライフサイクルの活動を支援する統合的特許構築システムについて報告する。

なお、統合的特許構築システムは、特許調査専門家だけでなく専門家並みの検索式を構築できる特許調査支援システム、特許請求の範囲等の発明を説明する文章を入力すれば、約 50% の特許出願書類を自動生成する特許出願書類半自動生成システム、および特許書類の品質を自動評価する特許書類品質評価システムからなる。

## 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、調査対象に関連したキーワードと、その同義語、上位語、下位語（以下、これらを関連語という）や、IPC、F ターム（以下、これらをコードという）などの特許分類コードで構成される検索式を用いる。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、特許公報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。

また、特許明細書作成フェーズを支援するツールとして、「PatentCreator」という特許明細書作成支援ツールがある。本ツールは、段落番号を自動付与する機能、作成した明細書から符号の説明を自動生成する機能などを有する。その他、本ツールに類するツールは、いくつか存在する。ただし、再利用性のある文をデータベース化して利用する等、特許書類作成の業務の効率を大幅に向上させるツールは見当たらない。

さらに、StraVision などの特許価値評価システムが存在するが、特許書類の解析が不十分であり、特許書類の品質評価としては利用できない。

## 3 統合的特許構築システムの概要

統合的特許構築システムは、特許検索システム、特許書類半自動生成システム（PatentGenerator）、および特許書類品質評価システム（PatentValueAnalyst）が連携するシステムである。つまり、ユーザが発明を行った後、特許検索システムで関連特許を検索し、関連特許を抽出する。そして、関連特許の特許書類から、要約の情報、効果の情報を取得し、特許書類半自動生成システムに与える。特許書類半自動生成システムは関連特許の情報を、明細書の【背景技術】に貼り付ける。そして、特許書類半自動生成システムにより、効率的に明細書が作成された後、特許書類品質評価システムにより特許書類の品質評価を行い、品質が所定以上になるまで、特許書類半自動生成システムを繰り返し利用して、特許書類を完成させる。なお、ここでの特許検索システムは、後述する特許調査支援システム（PatentSearchAssistant）を含み、PatentSearchAssistant が提案した検索式で、特許公報を検索するシステムである。

## 4 特許調査支援システム (PatentSearchAssistant)

### 4.1 PatentSearchAssistant の概要

PatentSearchAssistant 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、関連語辞書構築機能と検索式構築のための機能の 2 つの機能により構成される。システム全体の概要を図 1 に示す。

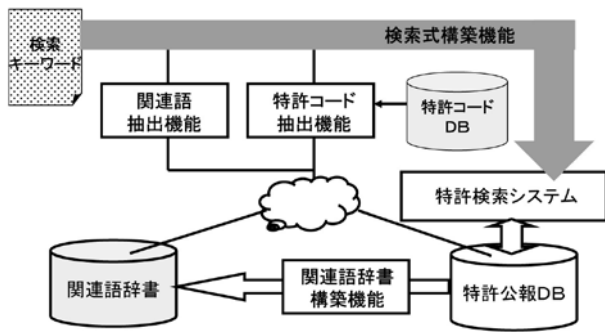


図1 PatentSearchAssistant の概要

## 4.2 関連語辞書構築機能

関連語辞書構築機能は、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、関連語辞書を構築する(2)。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語、下位語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」のように上位語、下位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。そして、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。

## 4.3 検索式構築のための機能

本機能は、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、適切な特許分類コード(以下、「コード」とも言う。)を抽出する特許コード抽出機能で構成される。

関連語抽出機能は、キーワードに対応した関連語を関連語辞書から取得する。取得する関連語は、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、ユーザが選択した関連語である。

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、検索に適したコードを取得する機能である。本機能において、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、検索語が出現する特許公報が保持する各コードにその出現割合を重みとして与え、コードごとに重みの総和(以下、コード重要度という)を算出する。そして、本機能において、コード重要度が閾値より高いコードを出力する。

出力する検索式は、取得したコードのすべての組み合わせと検索語から構成される。そして、ユーザは使用する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ものとした。

## 4.4 PatentSearchAssistant 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 + 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)」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 特許書類半自動生成システム (PatentGenerator)

## 5.1 PatentGenerator の概要

PatentGenerator は、明細書設計書を入力とし、特

許明細書、特許請求の範囲、および要約書を半自動生成するシステムである。明細書設計書とは、権利化したい発明の内容を記載した文書であり、通常、特許請求の範囲に加えて、「技術開示をする実施の形態の番号」、「発明の概要の説明」、「各請求項の文言の詳細な説明」などを記載する。「技術開示をする実施の形態の番号」とは、各請求項に記載の発明の技術開示を行う実施の形態の番号であり、請求項ごとに記載する。「発明の概要の説明」とは、請求項に記載の発明の概要を説明する文章であり、請求項ごとに記載する。「各請求項の文言の詳細な説明」とは、請求項内の用語の詳細説明である。なお、明細書設計書は、PatentGeneratorが予め想定した形式以外の形式で記載したとしても、その明細書設計書に対応した特許明細書等が生成されるが、予め想定した形式で記載する方が、より質の高い特許明細書が生成できる。

PatentGeneratorは、明細書設計書解析機能、特許書類生成機能、特許部品データベース構築機能を有する(図2参照)。明細書設計書解析機能は、明細書設計書を自然言語処理し、技術用語、構成要素名、および構成要素説明などを取得し、XMLのタグ付けを行う機能である。特許書類生成機能は、明細書設計書解析機能の出力を用いて、3種類の特許部品DBを検索し、特許庁フォーマットに従って、特許書類を生成する機能である。また、特許部品データベース構築機能は、特許部品データベース構築機能は、特許部品DBを自動構築する機能である。

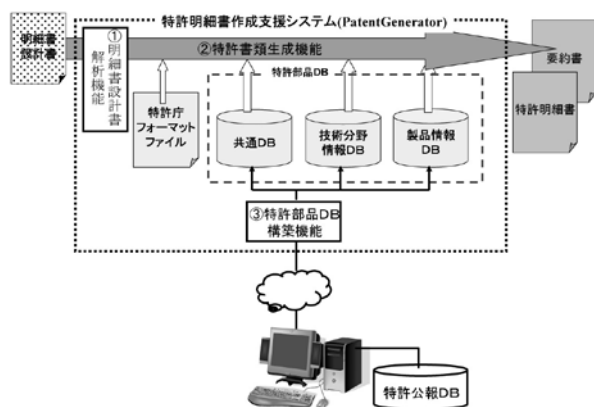


図2 PatentGeneratorの概要

## 5.2 特許部品DB

特許部品DBには、共通情報DB、技術分野情報DB、製品情報DBがある。

共通情報DBは、特許庁フォーマットの各見出しに対応する定型文を格納している。

技術分野情報DBは、技術分野(電気、機械、化学、コンピュータ・ソフトウェアなど)ごとに構築されるDBであり、その技術分野の技術用語の説明文や、その技術分野の発明の権利強化を図るための定型文や、技術分野固有の発明の説明手順の情報の集合である。定型文は、実施可能要件(特許法第36条第4項)を満たす意義や、特許の権利範囲を拡大する意義を有し、特許明細書作成におけるノウハウを表現した文章である。

製品情報DBは、発明の対象の製品ごとに構築されるDBであり、発明を構成する構成要素の実現手段の情報の集合である。

## 5.3 特許書類生成の手順

PatentGeneratorは明細書設計書解析機能と、特許書類生成機能により、特許書類を生成する。明細書設計書解析機能は、明細書設計書を自然言語処理し、特許請求の範囲特有の手がかり句を用いて、発明を構成する構成要素名、技術用語を取得する。特許書類生成機能は、明細書設計書解析機能が取得した構成要素名、技術用語をキーとして、特許部品DBを検索し、構成要素や技術用語の説明文を取得し、特許明細書の一部、特許請求の範囲、要約書を自動的に生成する。また、特許書類生成機能は、生成できなかった特許明細書の該当箇所に、特許明細書の記載に有効なガイドも出力する。

## 5.4 特許部品データベース構築機能

特許部品データベース構築機能は、特許部品DBを自動構築する機能であり、以下の動作をする。まず、特許公開公報の【符号の説明】から構成要素名(例えば、「情報格納手段」)を取得し、この構成要素名を形態素解析し、形態素に分割する(例えば、「情報|格納|手段」)。そして、「手段」「部」などの手がかり句を除

いた最後の形態素（例えば、「格納」）および構成要素名を取得する。そして、最後の形態素および構成要素名をキーとして特許公報を検索し、「〈構成要素名〉は、～である。」という文（例えば、「情報格納手段は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。」）を取得し、構成要素名の箇所（例えば、「情報格納手段」）を、変数〈構成要素〉に置き換え、「〈構成要素〉は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。」を得る。

また、複数の特許公報の同一のタグ領域を比較し、同一または類似度が一定以上の文を抽出し、共通DBのレコードとする。さらに、特許公報から専門用語を抽出し、定義文の手がかり情報（例えば、「〈専門用語〉とは、～である。」）を用いて専門用語の定義文を抽出し、技術分野情報DBのレコードとする。ただし、各特許部品DBに登録する際には、予め作成した不要パターン（例えば、「構成要素の説明に構成要素を使用している文章」）を用いて再利用性の低い文章が登録されないようにしている。

## 5.5 PatentGenerator の評価

PatentGenerator を利用したユーザの評価によれば、特許出願書類の作成効率は 1.8 倍向上した。また、PatentGenerator を用いた場合、特許明細書の品質が大幅に向上した。また、特許明細書の標準化が図られ、第三者による特許明細書の品質チェック時間が大幅に削減された。

# 6 特許書類半品質評価システム (PatentValueAnalyst)

## 6.1 PatentValueAnalyst の概要

PatentValueAnalyst は、人手による特許の評価結果が格納された教師データベースを利用し、評価対象の特許書類を特許書類解析機能により得られた約 100 の

3 参照)。

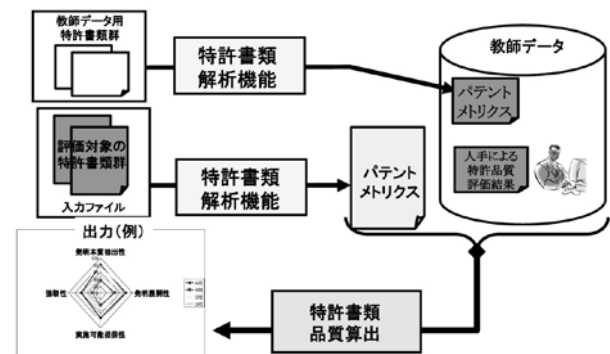


図3 特許書類品質評価システムの概要

## 6.2 パテントメトリクス

パテントメトリクスとは、特許書類の品質に影響すると考えられる数値パラメータのことである。

パテントメトリクスのうち、代表的なものとしては、「独立形式請求項の数」、「ネストレベル」、「実施の形態において説明されていない構成要素の数」、「実施の形態において説明されていない用語の数」、「不適切な引用形式請求項の数」、「主語のない文の数」などがある。なお、「ネストレベル」とは、請求項間の引用関係の深さである。

「独立形式請求項の数」は、請求項中の「請求項2記載の」などの引用関係を示す文言を手がかり句とし、独立形式請求項か、引用形式請求項かを判断することで算出する。また、「ネストレベル」は、請求項間の引用関係を解析し、請求項間の2項関係を取得し、それを基に、請求項のツリー構造を構築し、その深さを算出する。また、「実施の形態において説明されていない構成要素の数」は、請求項から手がかり句を基に構成要素名を取得し、この取得した構成要素名が、実施の形態において出現するか否かを判断することで算出する。また、「実施の形態において説明されていない用語の数」は、請求項を形態素解析し、名詞の連続である名詞句を用語として取得し、この取得した用語が、実施の形態において出現するか否かを判断することで算出する。また、「不適切な引用形式請求項の数」は、「ネストレベル」の算出と同様に、請求項間の引用関係を解析し、請求項

間の2項関係を取得し、存在していない請求項を引用していないか、自身の番号よりも大きい番号の請求項を引用していないかなどを判断し算出する。また、「主語のない文の数」は、実施の形態の文章を文ごとに分割し、その文ごとに形態素解析を行い、主語となり得る品詞が存在するか否かを判断し、算出する。

### 6.3 評価方法

本システムにおいて、特許書類品質の評価者は、約100の Patent Metrics を考慮して、特許書類の品質を評価している、と仮定している。

この仮定を基に、教師データベースに、特許書類から取得した Patent Metrics と、特許書類の評価結果とを対にして、所定数以上を格納している。そして、評価したい特許書類を本システムに入力すれば、特許書類解析機能により、評価対象の特許書類から約100の Patent Metrics を抽出し、それらの Patent Metrics と教師データベースから、機械学習機能を用いて、評価対象の特許書類の品質を自動算出する。なお、本システムで用いている機械学習機能は、サポートベクター回帰 (Support Vector Regression : SVR) と言われるアルゴリズムに基づく。

### 6.4 本システムの評価結果

20件の教師データを用意し、5件の評価対象の特許書類を入力し、本システムを評価した。評価は、100点満点の特許明細書の品質評価である。評価結果を表1に示す。評価結果によれば、本システムでの評価は、人手による評価と比較して、誤差は出るが、人手による評価の傾向は概ね踏襲していることが確認できた。

|      | 人手評価 | 自動評価 |
|------|------|------|
| 特許 1 | 32   | 42   |
| 特許 2 | 74   | 64   |
| 特許 3 | 20   | 30   |
| 特許 4 | 30   | 40   |
| 特許 5 | 40   | 54   |

表1 本システムの評価結果

## 7 おわりに

以上、特許調査と特許書類作成を有機的に支援する統合的特許構築システムについて説明した。

さらに、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、特許書類の作成効率と品質を大幅に向上する特許出願書類半自動生成システムおよび特許書類品質評価システムの機能およびそれらの効果について説明した。

なお、今後、特許工学の研究を進め、統合的特許構築システムをさらに拡張し、さらに多数のツールが連携するシステムの開発を行っていきたい。

### 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.