

特許明細書半自動生成システムと 特許部品データベース構築技術

IRD国際特許事務所所長／弁理士
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、博士(情報学)。弁理士会、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／技術用日本語プラットフォーム委員会委員

✉ | htanigawa@ird-pat.com

☎ | 06-6944-4530

1 はじめに

従来、弁理士や企業の知財部門、研究開発部門においては、個人のスキルに頼った特許開発が行われていたため、高い生産性で特許明細書が作成されているとはいえなかった。また、特許明細書の品質の維持は、もっぱら個人のスキルに頼っており、品質管理の概念が浸透していないのが現状である。

そこで、本論文では、特許開発を効率化、高品質化するために、3種類の特許部品データベースを再帰的に検索し、特許明細書を構成する文章を取得し、特許庁フォーマットに従って、取得した文章を合成することにより、特許明細書を生成する特許明細書半自動生成システム(Patent Generator「以下、適宜、「PG」ともいう。)」について述べる。

本稿の構成は以下のとおりである。2章では、関連研究・技術について述べる。3章では、特許部品データベースに基づく特許明細書生成エンジン(PG)について述べる。4章では、PGを用いて、コンピュータ・ソフトウェア関連発明の特許明細書を作成した際の評価結果について述べる。5章では、PGの性能に重要な役割を果たす特許部品データベースを自動構築する技術について述べる。最後に6章では結論について述べる。

2 関連研究・技術

特許明細書の作成を支援するツールとして、PATENTBOY¹⁾や、PATENTEDITOR²⁾などが存在する。

これらのツールは、特許出願に必要な特許明細書中の段落番号を自動付与したり、概ね完成した特許明細書から特許図面の符号の説明を自動生成したり、定型文書を複写したりする機能を有する。これらの既存ツールは、蓄えたノウハウや技術の実現手段についての文章をデータベース化して、利用するアプローチを採っていないので、特許明細書の品質向上には寄与せず、作成効率の向上も不十分である。

3 特許明細書半自動生成システム (Patent Generator)

3.1 Patent Generator (PG) の概要

PGは、発明を記載した文章を入力として、特許明細書を半自動生成するシステムである。発明を記載した文章の中でも、特許請求の範囲や明細書設計書(以下、明細書設計書等という。)という、一定の形式を有した文を入力とした場合、PGは、非常に有用な特許明細書の

一部分を出力できる。

明細書設計書は、書類「特許請求の範囲」に加えて、特許明細書の設計のために3種類の情報を記載したドキュメントであり、特許明細書を記載するための指針となる成果物である。明細書設計書の例を、図1に示す。明細書設計書には、特許請求の範囲に加えて、以下の3種類の情報を記載する。

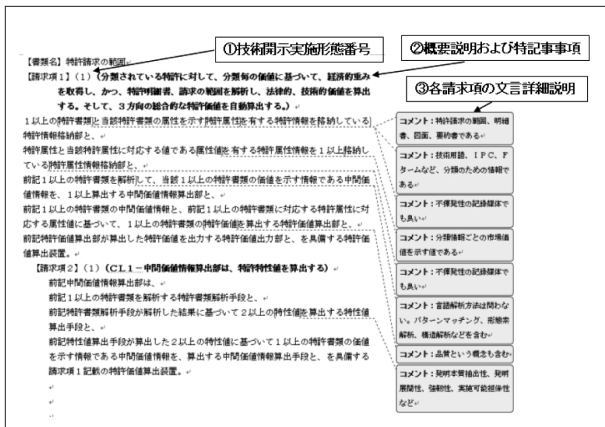


図1 明細書設計書の例

・技術開示実施形態番号

技術開示実施形態番号は、各請求項の技術開示を行う実施の形態の番号であり、請求項ごとに記載される。

・概要説明および特記事項

概要説明は、請求項の概要である。請求項の概要説明が実施の形態の概要となる。また、特記事項は、各請求項の技術開示を行うにあたって、特に触れたい事項である。特記事項は、請求項ごとに記載されるが、必須の項目ではない。

・各請求項の文言詳細説明

請求項の文章中の用語の中で、特に、説明を要する用語が存在する。文言詳細説明は、その用語に対する説明である。説明を要する用語は、例えば、発明の説明のために用いた造語である。文言詳細説明は、図3の「コメント」である。

また、PGは、PGデータベース、およびPG処理部からなる。そして、PGデータベースは、特許庁フォーマット、と3種類の特許部品データベースからなる（図2参照）。

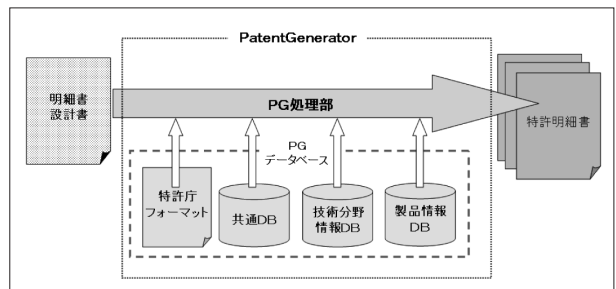


図2 PGの構造

特許庁フォーマットは、特許明細書、特許請求の範囲などの特許出願に必須の特許書類の記載項目を定義している。また、3種類の特許部品データベースとは、技術分野や製品の種類に依存しない文章等が格納された共通データベース、発明の技術分野に依存する文章等が格納された技術分野情報データベース、製品固有の文章を格納している製品情報データベースである。

さらに、PG処理部は、明細書設計書を自然言語処理し、発明を構成する構成要素名、技術用語などを取得する。そして、構成要素名、技術用語などを用いて、PGデータベースを検索し、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、PGは、特許明細書を記載する指針となる適切なガイドをも出力する。

PGは、明細書設計書に対して、形態素解析、係り受け解析を行い、精度高く構成要素名、および構成要素の階層関係を取得する。そして、取得した構成要素名、および構成要素の階層関係を修正できるエディタ（図3参照）を用意して、特許明細書の生成に完全を期している。高度な自然言語処理技術を適用して、構成要素名、および構成要素の階層関係の取得の精度が大幅に向上したが、明細書設計書の入力のゆらぎに完全に対応することは困



難だからである。PGの出力例を図4に示す。

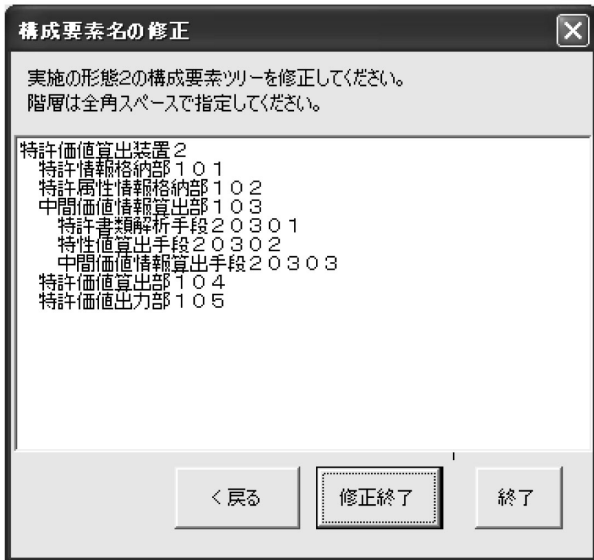


図3 構成要素の編集エディタ

広い範囲の技術開発を行い、毎年、多数の特許出願を行う大企業でPGを利用する場合、共通データベースと、技術分野情報データベースと、技術分野情報データベースの数の比は、1:n:m (1<n<=m) である。

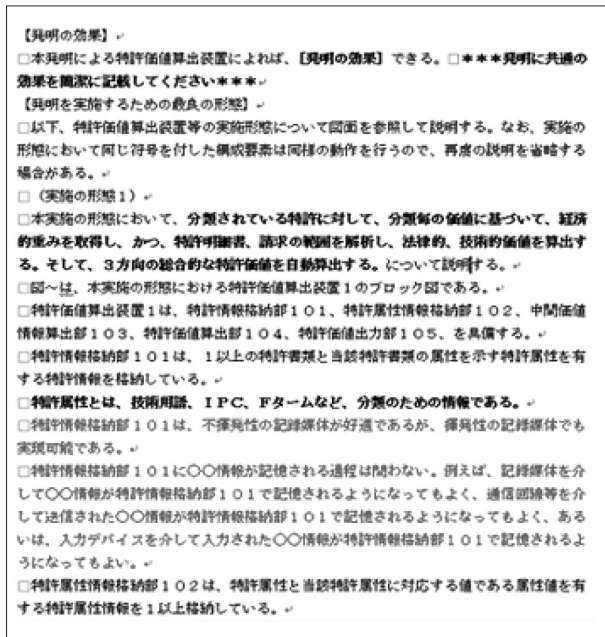


図4 PatentGeneratorの出力明細書の例

そして、このような企業において、多数の事業部門を統括する部署で共通データベースを構築し、各事業部門で技術分野情報データベースを構築し、各製品担当部門で技術分野情報データベースを構築する、という具合に、3種類のデータベースを分散開発し、かつ、特許明細書の作成者は、PGを用いて、統合的に3種類のデータベースを利用する。

3.2 PGデータベース

PGデータベースを構成する3種類の特許部品データベースについて述べる。

(1) 共通データベース

共通データベースは、特許庁フォーマットのタグに対応する定型文、または説明手順の情報を格納している。説明手順の情報とは、発明を説明する手順を定義する情報である。共通データベースの属性は、「タグ」および「タグ説明情報」である。「タグ」は、特許庁フォーマットで規定されているものであり、墨付き括弧 (【】) で示される。「タグ説明情報」は、定型文や、説明手順を定義する情報が属性値となる。「タグ説明情報」は、予約語変数を含む。予約語変数は、PGで予め決められている変数であり、例えば、「発明の名称」や「構成要素」などである。予約語変数の値は、明細書設計書を解析して取得される。

(2) 技術分野情報データベース

技術分野情報データベースは、技術分野（電気、機械、化学、コンピュータ・ソフトウェアなど）ごとに構築されるデータベースであり、その技術分野の発明の権利強化を図るための定型文や、文献としての役割を果たすための定型文や、技術分野固有の発明の説明手順の情報の集合である。技術分野情報データベースは、「説明項目」「定型文・説明手順」の属性を有する。「説明項目」は、主として技術分野固有の説明するべき項目名である。また、「定型文・説明手順」は、主として、「説明項目」に

対応する定型文や、説明手順の情報である。定型文は、特許明細書の明瞭性や論理性を担保し、十分に技術開示を行い、発明が実施可能であるように記載しなければならない、という実施可能要件（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。

(3) 製品情報データベース

製品情報データベースは、発明の対象の製品ごとに構築されるデータベースであり、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。例えば、コンピュータ・ソフトウェア関連発明の特許明細書において、記載しなければならない実現手段は、コンピュータおよび周辺機器を構成する部品の実現手段である。コンピュータおよび周辺機器の部品は、CPUや記録媒体やディスプレイなど、概ね決まっている。製品情報データベース中の情報は、権利範囲を拡大するための実現手段、または発明の成立性（特許法第29条第1項柱書）や実施可能要件（特許法第36条第4項）といった特許要件を満たすための実現手段を定義したレコードである。なお、部品の実現手段は、発明の成立性（特許法第29条第1項柱書）や実施可能要件（特許法第36条第4項）の特許要件を満たすために、通常、記載しなければならない。さらに、実現手段は製品により異なるので、例えば、発明の対象の製品がテレビであれば、テレビ用の製品情報データベースが存在し、エアコンであれば、エアコン用の製品情報データベースが存在することとなる。

3.3 Patent Generatorの動作概要

PGを構成するPG処理部は、明細書設計書等に記載された各請求項を自然言語解析し、請求項ごとに、発明の名称、1以上の構成要素名、各構成要素の説明などの予約語変数の値、構成要素間の階層関係、および請求項数などの予約変数の値を取得する。

そして、PG処理部は、請求項ごとに、明細書設計書

等に記載された「技術開示実施形態番号」、「概要説明および特記事項」、「文言詳細説明」の3種類の情報を取得する。そして、取得した「概要説明および特記事項」は、予約語変数<概要説明および特記事項>の値となる。なお、「技術開示実施形態番号」および「文言詳細説明」は、実施の形態における構成要素説明の文章生成に利用される。

次に、PG処理部は、特許庁フォーマットを読み出す。そして、特許庁フォーマットに記載されたタグ（**[I]**）をキーとして、共通データベースからレコードを検索し、検索したレコードのタグ説明情報を、タグの直後に追記する。そして、共通データベースのタグ説明情報が予約語変数を含む場合、既に得ている予約語変数の値をタグ説明情報に代入し、タグの直後に挿入する文章を生成する。

次に、共通データベースのタグ説明情報が一般変数を含む場合、技術分野情報データベースの検索処理を行い、一般変数の値を取得する。この処理を、特許庁フォーマットに含まれるタグ（**[I]**）の数だけ繰り返す。

4 PGの評価

4.1 PGを利用した場合の作業効率向上

3名の特許明細書作成者が、約1年間にわたり、PGのプロトタイプ版を使用し、約100件の特許明細書を作成した。PGを利用して作成した特許明細書の技術分野は、コンピュータ・ソフトウェア関連発明である。PGを利用した場合、特許明細書作成に要した作業時間は、約15%減少した。また、PGを用いて作成した特許明細書のうち、10件の特許明細書を取り出して、PGによる文章の生成率を算出した。その結果、平均39.5%の情報、PGにより自動生成された。PGの文章の生成率



が、そのまま作業時間の短縮に結びついていないのは、定型的な文章や多数の特許明細書で共用される文章の作成が、発明特有の文章の作成ほど、労力がかからないからである、と考えられる。ただし、PGを用いて特許明細書を作成する場合、発明特有の文章の作成に注力すれば良いので、特許明細書作成の負担は、時間や量の効率化以上に減じられていると考える。

また、PatentGeneratorを利用したユーザ（年間、数件の特許書類を作成する研究者、技術者）の評価によれば、特許明細書の作成効率は1.8倍向上した。

4.2 PGによる特許明細書の品質向上

PGを用いて作成した特許明細書10件と、PGを用いずに作成した特許明細書10件の特許明細書品質を独自開発の特許明細書品質評価ツール³⁾を用いて算出した。PGを用いずに作成した特許明細書10件は、コンピュータ・ソフトウェア関連発明の分野の特許明細書であり、公開特許公報からランダムに抽出した。その評価結果を図5に示す。評価結果によれば、PG利用の特許明細書品質がPG非利用の特許明細書品質を大きく上回った。PGを利用して作成した特許明細書とPGを利用していない特許明細書とを比較して、特に、実施可能担保度と発明展開度において、高い数値が出ている。また、特許

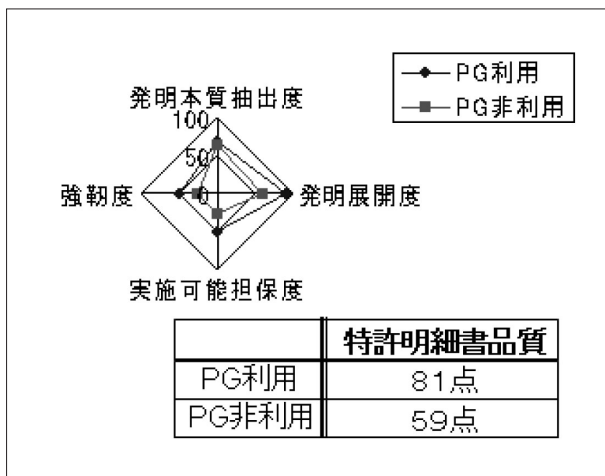


図5 PGによる特許明細書の品質向上

明細書の作成に熟練した弁理士などのノウハウをPGデータベースに登録することにより、特許明細書作成の経験の浅い者でも、PGを利用して特許明細書を作成すればその品質は、一定以上になるものと期待できる。

5 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、複数の特許公開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。

また、特許公開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。

さらに、特許部品データベース構築機能は、以下の動作をし、製品情報DBのレコードを取得する。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」）を取得する。そして、「＜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、特許公開公報から取得し、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。

6 おわりに

本論において、高品質な特許明細書を効率的に作成するための特許明細書半自動生成システム（Patent Generator）について述べた。

さらに、Patent Generatorにより、特許明細書の作成効率が大幅に向上したこと、および特許明細書の品質向上が図られたと考えられることについても報告した。

参考文献

- 1) URL [http://www.cks.co.jp/html/m-2.htm]
- 2) URL [http://www.hypertech.co.jp/products/pc/index.html]
- 3) 谷川英和、新森昭宏、大屋由香里、奥村学、難波英嗣「特許明細書の作成・分析・評価を支援する統合的特許工学支援システム」2007年度 日本知財学会。

