

特許記述言語 (PML) と 統合的特許工学システム

IRD 国際特許事務所 所長 / 弁理士
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。博士(情報学)、弁理士会、情報処理学会各会員。2007年度から特許産業日本語委員会委員。

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

IRD 国際特許事務所
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科学研究科修了。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)。



株式会社インテックシステム研究所 ICT 研究部長

新森 昭宏

PROFILE

1983年京都大学理学部卒業。1990年コロラド大学コンピュータサイエンス科修士課程修了。2005年東京工業大学総合理工学研究科博士課程修了。1983年(株)インテック入社。現在、(株)インテックシステム研究所 ICT 研究部長。博士(工学)。技術士(情報工学部門)。



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、特許ライフサイクルにおける各種作業を支援する既存ツールは多数、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。

また、高度な特許検索を実現するツールが既に存在しているが、特許公開公報をはじめとする特許情報の構造化が不十分なため、ツールが提供する機能にも限界があった。

そこで、本論文において、統合的な特許工学システ

ムの実現、および高度な機能を有するツールの実現等のために、特許書類を精緻に構造化する特許記述言語(PML)と、統合的特許工学システムについて報告する。また、統合的特許工学システムを構成する特許明細書半自動生成システムについて述べる。

2 関連研究・技術

特許庁指定のXMLフォーマットが存在する。しかし、各種ツールが高度な機能を発揮するためには、特許庁指定のXMLフォーマットでは、構造化の粒度が大きい。

また、請求項を構造化した特許請求項記述言語(PCML)が存在する²⁾。PCMLは、特許請求の範囲について、構造化が進んでいるが、明細書等には及んでいない。したがって、各種ツールが高度な機能を発揮するには、不十分である。なお、特許請求項記述言語(PCML)は、2003年10月以降活動されている知的財産権関連の学術研究会である「特許戦略工学分科会」で研究、提案されている言語である。

3 特許記述言語 (PML)

(1) PML のコンセプト

PML (Patent Markup Language) は、主として、特許請求の範囲、明細書、要約書から、特許情報を多面的に構造化するマークアップ言語である。ただし、PML は、明細書等の出願書類以外の特許に関する広範囲な情報（審査書類情報など）から得られる情報をも用いて、特許情報をマークアップする言語である。なお、ここでは、特許情報とは、特許に関するあらゆる情報を総称する用語として用いる。

PML は、各種のツールがより高度な機能を提供できるようにすることを第一の目的としている。さらに、PML は、複数のツールが連携して、高度な機能を提供できるようにすることを第二の目的としている。

このような目的を達成するために、PML は、意味のある単位で、できるだけ細かく特許情報を構造化する、というコンセプトを有する。

また、PML の設計にあたって、特許工学ツールの特

性が考慮されている。つまり、特許工学ツールには、業務推進系ツール、管理系ツール、および分析評価系ツールが存在する。そして、業務推進系ツールは、特許請求の範囲、明細書などの特許情報の内容や構造を利用する機会が多い。また、管理系ツールは、特許請求の範囲、明細書などの特許情報の内容を利用する機会が多い。さらに、分析評価系ツールは、特許情報の属性や構造を利用する機会が多い。

そこで、PML を、i) 内容 (text)、ii) 属性 (info)、iii) 構造 (structure) の3つの観点から、特許情報をタグ付けする言語とした。

具体的には、特許請求の範囲の各請求項には、請求項の文そのものに付される <claim-text> タグ、請求項の属性を示す情報に付される <claim-info> タグ、請求項の構造を特定する情報に付される <claim-structure> タグの3つのタグが出現する。そして、3つのタグの下位には、図1に例示するように、詳細な各種タグが付され、特許情報が精緻に構造化される。

また、各種特許工学ツールが PML を利用する場合、ツールにとって利用可能なタグのみを利用し、利用可能

```
<claims>
  <subject-of-claims>虹彩情報を用いて、入退室管理を行う</subject-of-claims>
  <effect-of-claims>入退室管理が容易になる</effect-of-claims>
  <purpose-of-claims>入退室管理を強化すること</purpose-of-claims>
  <claim num="1">
    <claim-text>セキュリティの施された建屋または施設の出入口近傍に設置され且つ入退室者の職から虹彩情報を入力する虹彩情報入力手段と、前記出入口より入退室する入退室者の入退室者数を計数する入退室者計数手段と、前記虹彩情報入力手段が入手した虹彩情報と予めデータベースに登録された虹彩情報とを照合し、前記虹彩情報入力手段が入手した虹彩情報と前記入退室者計数手段が計数した入退室者数とを照合してこれらが一致しているときは正常情報を、一致していないときは異常情報を出力する入退室管理装置を備えたことを特徴とする入退室管理システム。</claim-text>
    <comment id="claim_1_1">
      <target-of-explanation>建屋</target-of-explanation>
      <explanation>ビルでも、一軒家でも良い。</explanation>
    </comment>
    <comment id="claim_1_2">
      <target-of-explanation>出入口近傍</target-of-explanation>
      <explanation>出入口が見える範囲であれば良い。好ましくは、出入口から3m以内</explanation>
    </comment>
    <effect-of-claim>虹彩情報を用いて、入退室管理が容易になる</effect-of-claim>
    <claim-info claim-title="入退室管理システム" number-of-substantial-defining-matters="3" type1="Independent" type2="elementEnumeration" />
    <claim-structure>
      <body-part>
        <defining-matter level="1" seqno="1" type="substantial" stype="component">
          <segment level="2" seqno="1" stype="modifier">
            <phrase id="1_1" modifying-phrase-id="1_2">セキュリティの</phrase>
            <phrase id="1_2" modifying-phrase-id="1_3">施された</phrase>
            <phrase id="1_3" modifying-phrase-id="1_5">建屋</phrase>
            <phrase>または</phrase>
            <phrase id="1_4" modifying-phrase-id="1_5">施設の</phrase>
            <phrase id="1_5" modifying-phrase-id="1_6">出入口近傍に</phrase>
            <phrase id="1_6" modifying-phrase-id="1_7">設置され</phrase>
            <phrase id="1_7" modifying-phrase-id="0">且つ</phrase>
            <phrase id="1_8">入退室者の</phrase>
            <phrase id="1_9">職から</phrase>
            <phrase id="1_10">虹彩情報を</phrase>
            <phrase id="1_11" modifying-phrase-id="0">入手する</phrase>
          </segment>
          <segment level="2" seqno="2" stype="element" id="claim_1_body_1122">虹彩情報入力手段</segment>
          <segment level="2" seqno="3" stype="postpositional">と、</segment>
        </defining-matter>
        <defining-matter level="1" seqno="2" type="substantial" stype="component">
          <segment level="2" seqno="1" stype="modifier">
            <phrase>前記出入口より</phrase>
            <phrase>入退室する</phrase>
            <phrase>入退室者の</phrase>
            <phrase>入退室者数を</phrase>
            <phrase>計数する</phrase>
          </segment>
          <segment level="2" seqno="2" stype="element" id="claim_1_body_1222">入退室者計数手段</segment>
          <segment level="2" seqno="3" stype="postpositional">と、</segment>
        </defining-matter>
      </body-part>
    </claim-structure>
  </claim num="1">
</claims>
```

図1 特許書類に PML を付与した例



でないタグは利用しない、ということをツールの設計思想とする。つまり、解釈不能なタグが存在する場合でも、ツールは、エラーを出さず、解釈不能なタグが付与された情報を無視することとなる。

さらに、PML において、特許庁の XML フォーマットに規定されているタグは、そのまま利用している。将来的な標準化を目指して、PML に、既に確定、運用されているタグを採り入れた。

(2) PML の内容

PML の DTD (Data Type Definition) において、特許請求の範囲、明細書、および要約書を、複数の要素 (ELEMENT) に分割している。また、DTD において、各要素の属性値は、属性 (ATTLIST) として定義している。属性は、「86」存在する。

図 1 に示したように、<claims> は特許請求の範囲を示すタグである。また、図 1 の <subject-of-claims> は特許請求の範囲全体の課題、<effect-of-claims> は特許請求の範囲全体の効果、<purpose-of-claims> は特許請求の範囲全体の目的を示す。また、<claim> 内の <claim-info> タグの属性値として、発明の名称を示す “claim-title”、発明特定事項の数を示す “number-of-defining-matters”、独立形式請求項か、引用形式請求項かを示す “type1”、ジェブソンの形式か、順次列挙形式か、構成要素列挙形式かを示す “type2” 等の属性値がある。さらに、<claim> 内の <claim-structure> の下位に、特徴部または請求

項全体を示す <body-part> が存在し、<body-part> の下位に、発明特定事項を表す <defining-matter> が存在する。そして、<defining-matter> はさらに <segment> に分割される。そして、<segment> はさらに <phrase> に分割される。

4 統合的特許工学システム

統合的特許工学システムは、PML をハブとして、多数のツールが連携するシステムである。

現在想定している、統合的特許工学システムの連携の例を図 2 に示す。図 2 は、以下のことを説明する。ユーザが発明を行った後、特許検索ツールで関連特許を検索し、関連特許の特許書類を PML 化する。そして、関連特許の特許書類の PML から、例えば、要約の情報、効果の情報を取得し、明細書作成支援ツールに与える。明細書作成支援ツール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、明細書作成支援ツールにより、効率的に明細書が作成された後、特許明細書品質評価ツール、および特許明細書読解支援ツールは、それぞれ必要な情報を取得し、特許明細書の品質評価、読解支援を行い、特許明細書の作成を支援する。そして、出願書類が完成した後、明細書作成支援ツールは PML を出力し、当該 PML を出願支援ツールが読み取り、特許出願する。そして、権利化の各種処理が行われ、特許権が発生したのち、出願支援ツールが出力した PML を

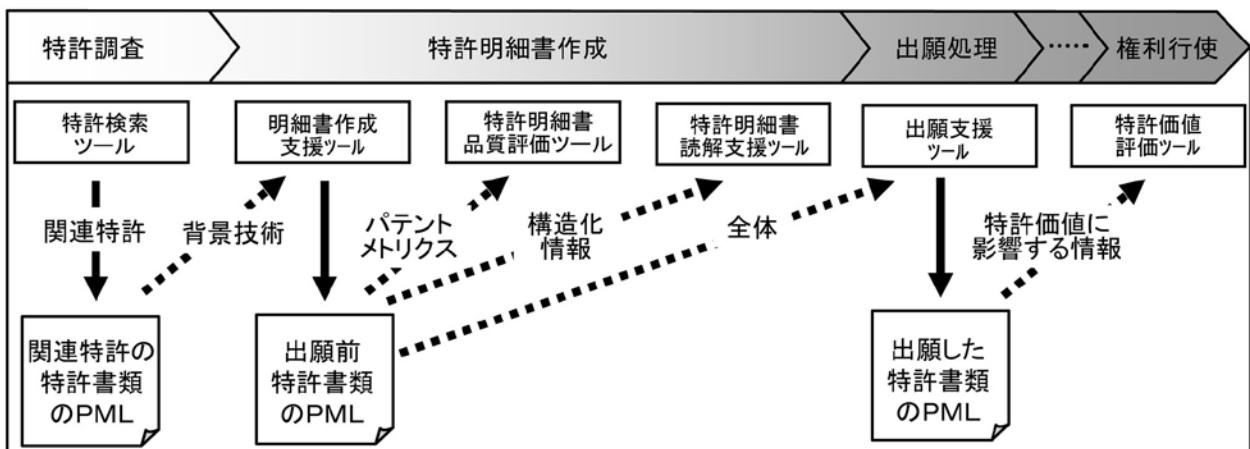


図 2 統合的特許工学システムの連携の例

用いて、特許価値評価ツールが特許価値を算出する。

5 特許明細書半自動生成システム (PatentGenerator)

以下、統合的特許工学システムを構成するツールの例である、PatentGenerator (以下、適宜「PG」とも言う。) について説明する。

(1) Patent Generator の仕組み

PatentGenerator は、明細書設計書を入力とし、特許明細書スタイルファイル、3種類の特許部品 DB を用いて、特許請求の範囲、特許明細書、要約書の3種類の書類を半自動生成するツールである。図3は、その概要を示したものである。

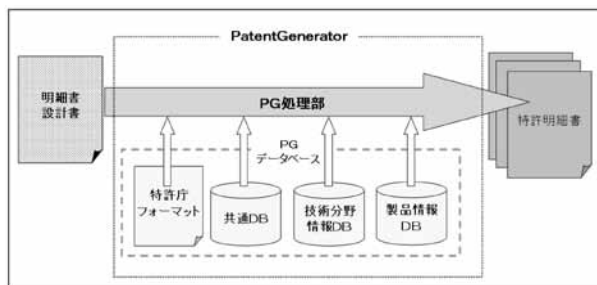


図3 PatentGenerator の概要

特許明細書スタイルファイルは、生成する特許明細書の書式・様式を定義したものであり、PatentGenerator は、明細書設計書の解析結果、および3種類の特許部品 DB の部品を特許明細書スタイルファイルの書式に成型しながら特許明細書を生成する。

(2) 明細書設計書

明細書設計書は、特許請求の範囲に加えて、権利化したい発明に関する情報を記載した文書である(図4参照)。具体的には、各請求項の技術開示を行う実施の形態番号、各請求項に係る発明の概要説明、各請求項中の技術用語の説明、各請求項に係る発明が解決する課題、各請求項に係る発明の効果、などを記載する。

(3) 特許部品 DB

特許部品 DB は、特許明細書作成に係るノウハウや、技術説明のための文章などを蓄積したものであり、製品情報 DB、技術分野情報 DB、共通情報 DB の3種類から構成される。

共通情報 DB は、同様の技術分野の発明であれば共通して使用することのできる定型文が蓄積される DB である。また、技術分野情報 DB は、任意の技術分野特有の専門用語の説明文が蓄積される DB である。さ

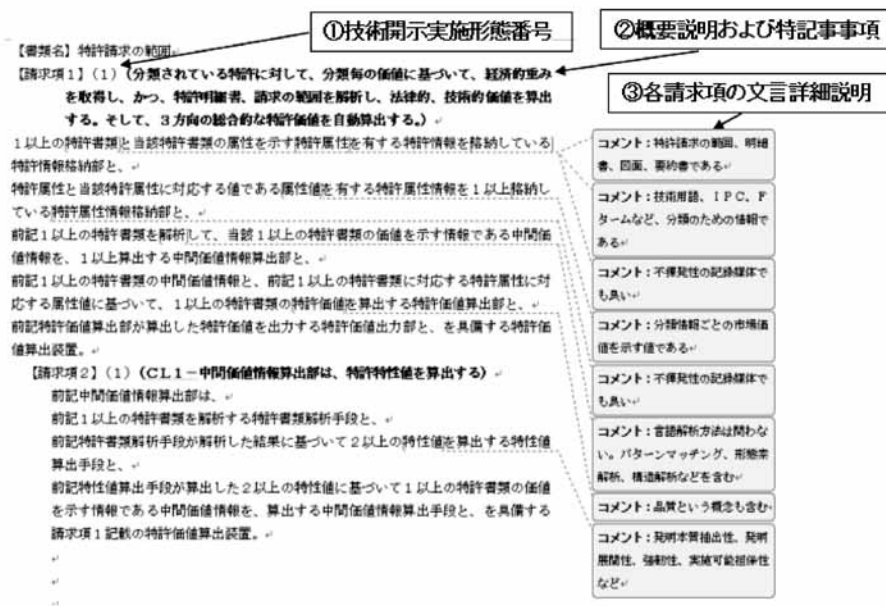


図4 明細書設計書の例



らに、製品情報 DB は、発明を構成する構成要素の実現手段の説明文が蓄積される DB である。

(4) 特許部品 DB 構築機能

PatentGenerator の特許部品 DB 構築機能は、ユーザが指定した特許公報を解析し、共通情報 DB、技術分野情報 DB、製品情報 DB の 3 種類の特許部品 DB を構築する機能である。図 5 は、その概要を示したものである。

共通情報 DB については、複数の特許公報における同一領域（墨付き括弧で囲まれた領域）の文章を、一文ずつ比較し、類似度が一定以上、および複数の特許公報間における出現頻度が一定以上ある文を取得することで構築する。例えば、上記の条件に一致する文として【発明を実施するための形態】付近から「以下、本発明の実施の形態について、図を参照しながら説明する。」が取得できた場合、取得できたタグの名称である「発明を実施するための形態」を「名称」の欄に、「以下、本発明の実施の形態について、図を参照しながら説明する。」を「ひな型」の欄に登録する。

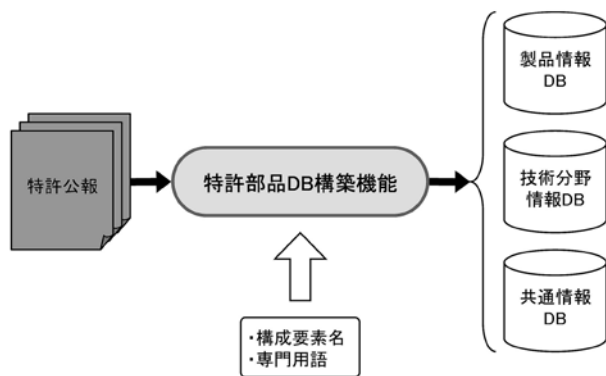


図 5 特許部品 DB 構築機能の概要

技術分野情報 DB については、明細書設計書の解析の結果得られた専門用語を基に、「<専門用語>とは、～である。」の形式の文の特許公報から取得することで構築する。例えば、明細書設計書から「ビットマップデータ」という専門用語が抽出でき、それを基に特許公報から「ビットマップデータとは、ドットまたはピクセルと呼ばれる点の集合で表現された画像データのことで

ある。」の文が取得できた場合、「ビットマップデータ」を「名称」の欄に、「ビットマップデータとは、ドットまたはピクセルと呼ばれる点の集合で表現された画像データのことで。」を「説明」の欄に登録する。

製品情報 DB については、明細書設計書の解析の結果得られた構成要素名を基に、「<構成要素名>は、～である。」や「<構成要素名>は、～であってもよい。」の形式の文の特許公報から取得することで構築する。例えば、明細書設計書から「ばね」という構成要素名が取得でき、それを基に特許公報から「ばねは、コイルばねが好適であるが、引きばねであってもよい。」の文が取得できた場合、「ばね」を「名称」の欄に、「ばねは、コイルばねが好適であるが、引きばねであってもよい。」を「実現手段」の欄に登録する。また、明細書設計書から「情報格納手段」という構成要素名が取得でき、それを基に特許公報から「情報格納手段は、不揮発性の記録媒体であってもよいが、揮発性の記録媒体であってもよい。」の文が取得できた場合、「〇〇格納手段」のような同様の構成要素名にも適用できるように「格納」を「名称」の欄に登録し、構成要素名の箇所を、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であってもよいが、揮発性の記録媒体であってもよい。」を「実現手段」の欄に登録するようにしてもよい。

6 PatentGenerator の評価

(1) PatentGenerator 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% の情報が、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。

(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6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。

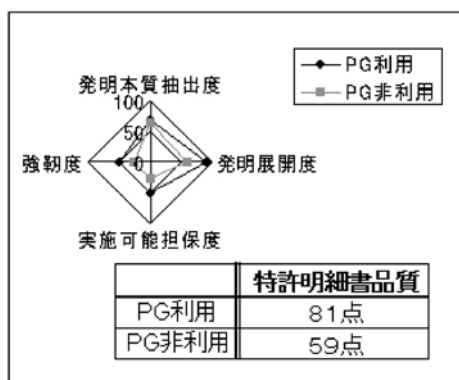


図6 PGによる特許明細書の品質向上

(3) 特許部品DB構築機能の評価

特許部品DB構築機能により、種々の分野の特許

構築基盤が短時間で構築できるものと考えられる。

定性的な評価であるが、著者らが作成した特許書類10件から、特許部品DB構築機能を用いてDBを構築したところ、概ね再利用性の高そうな文章が抽出できていた。

定量的な評価は今後の課題である。

7 おわりに

本論において、ツールがより高度な機能を提供できるようになること、および複数のツールが連携して、高度な機能を提供できるようになることを目的として、特許書類を精緻に構造化するマークアップ言語「PML」について報告した。また、PMLを用いて実現できる統合的特許工学システムについて述べた。さらに、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(PatentGenerator)について述べた。

なお、PMLの仕様は、下記のホームページに公開し、広く意見を求めている。

(URL)<http://www.ird-pat.com/pml.html>

謝辞：本論文は、独立行政法人情報通信研究機構(NICT)の平成20年度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) 谷川英和他, 特許工学入門, p1 ~ p7(2003), 中央経済社.
- 2) URL「http://www.patentisland.com/Patent_Strategy_Engineering/」
- 3) 谷川英和, 新森昭宏, 奥村学, 難波英嗣「パテントメトリクスによる特許品質評価および特許価値評価」2006年度JAPIO「特許情報活用の時代の検索と機械翻訳技術」, pp154-159, 2006.