

# 特許記述言語 PML を用いた統合的特許構築システム

## Integrated Patent Creating System by using Patent Markup Language

谷川英和, 渡辺俊規, 増満光 (以上、IRD 国際特許事務所, (有) アイ・アール・ディー)

新森昭宏, 高木慎也 (以上、(株) インテックシステム研究所)、難波英嗣 (広島市立大学大学院)

Hidekazu Tanigawa, Toshinori Watanabe, Hikaru Masumitsu(IRD Corp.), Akihiro Shinmori,

Shinya Takagi (INTEC Systems Institute, Inc.), Hidetsugu Nanba (Hiroshima City University)

### 概要

特許書類の特性を考慮した特許記述言語 (PML) をハブとして連携する特許検索、特許書類半自動生成、特許書類品質評価の 3 システムにより、特許調査から特許出願に至る作業を統合的に支援する統合的特許構築システムについて報告する。特許検索システムを構成する特許調査支援システムは、キーワードの入力だけで、特許公報で使用されている関連語と特許分類コードを含む高度な検索式を構築する。特許書類半自動生成システムは、発明の説明文から出願書類の約 50% を生成する。さらに、特許書類品質評価システムは、特許書類から自動抽出する約 100 種類のパテントメトリクスの値を素性として、機械学習により評価する。

### 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<sup>1)</sup>。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。

また、特許公開公報や特許登録公報や実用新案登録公報等 (以下、総称して「特許公報」という) には、非常に有用な知識・情報が豊富に含まれているが、特許公報の構造化が不十分なため、既存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、特許書類の特性を考慮した言語であり、特許書類 (特許請求の範囲、明細書、要約書を総称して「特許書類」という) を精緻に構造化する特許記述言語 (Patent Markup Language (PML)) をハブとして、有機的にシステムが連携することにより、統合的に特許構築の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。

また、統合的特許構築システム構成するシステムであり、約 15 年分 (約 500 万件) の特許公報から、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を用いて、特許調査専門家でなくても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 (PatentSearchAssistant) について説明する。

また、約 15 年分の特許公報から、法律を考慮した自然言語処理により抽出した特許部品 DB を用いて、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 (PatentGenerator)、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 (PatentValueAnalyst) についても述べる。

## 2. 関連業績

特許書類を構造化する技術として、特許庁指定のXMLフォーマットが存在する<sup>2)</sup>。しかし、特許庁指定のXMLフォーマットでは、各種ツールが高度な機能を発揮するには、構造化の粒度が大きく、構造化が不十分である。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<sup>3)</sup>。PCMLは、特許請求の範囲のみの構造化であり、明細書は構造化できない。

また、特許調査ツールとして、PATOLISなどの多数のツールがあり、特許調査の専門家は、通常、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タームなどの特許分類コード(以下、これらを「コード」という)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<sup>4)</sup>。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、特許価値評価ツールとして、StraVision<sup>5)</sup>がある。本ツールでは、特許書類や特許出願以降の経過を解析し、特許の価値を算出するツールであるが、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. 特許記述言語と統合的特許構築システム

### 3.1 特許記述言語(PML)のコンセプト

PML(Patent Markup Language)は、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。なお、ここでは、特許情報とは、特許書類や、特許の権利化までの経過に関する情報などを含み、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な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を考慮している。現在、市販されている特許工学ツールは、以下の3種類である。

- ・特許ライフサイクルの所定のフェーズにおける業務の推進支援を行う業務推進系ツール
- ・各フェーズにおける成果物や中間成果物等(例えば、特許明細書や特許情報)の管理を行う管理系ツール
- ・成果物や業務を分析したり、評価したりする分析評価系ツール

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と構造を使用する。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使用する。さらに、分析評価系ツールは、特許情報の属性と構造を使用する。

そこで、PMLを、内容、属性、構造の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、例えば、特許請求の範囲の各請求項には、請求項の文そのものを示す<claim-text>タグ、請求項の属性を示す<claim-info>タグ、請求項の構造を示す<claim-structure>タグの3つのタグを定義した。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。

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。

### 3.2 PMLの内容

PMLのDTD(Document Type Definition)において、特許書類を、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、98種類存在する。具体的な属性の例を表1に示す。表1は、属性名が出現する要素名(タグ名)、属性名、属性の説明という構造になっている。

また、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。<claim-text>は、請求項の本文を示す。また、<claim>内の<claim-info>の属

表 1：PML における属性の例

要素名	属性名	説明
claims	number-of-claims	請求項数
	number-of-independent-claims	独立形式請求項数
	number-of-categories	カテゴリー数
	nest-level	ネストレベル
claim-info	number-of-substantial-defining-matters	発明特定事項数
	number-of-characters	請求項の文字数
description	number-of-embodiments	実施の形態の数
	number-of-figures	図面数
	number-of-no-example-creation-terms	例示のない造語数
	number-of-no-subject-sentences	主語の無い文数

性として、発明の名称を示す「claim-title」、請求項が独立形式請求項か、引用形式請求項かを示す「type1」等の属性がある。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、特徴部または請求項全体を分割した<segment>が存在する。そして、<segment>はさらに<phrase>に分割される。

```
<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眼から虹彩情報を入力する虹彩情報入手手段と、前記出入口より入退室する入退室者の人数を計数する入退室者計数手段と、前記虹彩情報入手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、前記虹彩情報入手手段が入手した虹彩情報入手者数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には正常情報を、一致していないときには異常情報を出力する入退室管理装置とを備えたことを特徴とする入退室管理システム。</claim-text>
    <claim-info claim-title="入退室管理システム" type1="independent" type2="elementEnumeration"/>
    <claim-structure>
      <body-part>
        <defining-matter level="1" seqno="1" stype="component" type="substantial">
          <segment level="2" seqno="1" stype="modifier">
            <segment level="3" seqno="1" stype="sequence">
              <phrase>セキュリティの</phrase>
              <phrase>施された</phrase>
              <phrase>建屋</phrase>
              <phrase>または</phrase>
              <phrase>施設の</phrase>
              <phrase>出入口近傍に</phrase>
              <phrase>設置され</phrase>
            </segment>
            <segment level="3" seqno="2" stype="sequence">
              <phrase>且つ</phrase>
              <phrase>入退室者の</phrase>
              <phrase>眼から</phrase>
              <phrase>虹彩情報を</phrase>
              <phrase>入手する</phrase>
            </segment>
          </segment>
          <segment level="2" seqno="2" stype="element">虹彩情報入手手段</segment>
          <segment level="2" seqno="3" stype="postpositional">と</segment>
        </defining-matter>
        <defining-matter level="1" seqno="2" stype="component" type="substantial">
          <segment level="2" seqno="1" stype="modifier">
            <segment level="3" seqno="1" stype="sequence">
              <phrase>前記出入口より</phrase>
              <phrase>入退室する</phrase>
              <phrase>入退室者の</phrase>
              <phrase>人数を</phrase>
              <phrase>計数する</phrase>
            </segment>
          </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

### 3.3 統合的特許構築システムと PML の利用例

統合的特許構築システムは、PML を軸に、特許検索システム、特許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、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、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

### 4.1 特許調査支援システムの開発背景

調査者は、膨大な数の特許公報から調査対象となる特許公報を検索する。検索時には、調査対象に関連したキーワードだけではなく、調査対象を適切に取得するために、キーワードの関連語、およびコードで構成される検索式を用いることが一般的である。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、かつ適切な検索式でなければ有効な検索ができない。

つまり、特許公報では、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は適切な用語を用いることができない。

また、特許検索において、適合率の高い検索結果を得るためには、適切なコードを用いる必要がある。しかし、IPC は約 7 万種類、F タームは約 36 万種類も存在するため、経験の浅い調査者では適切なコードを選択できない。

そこで、キーワードを入力するだけで、適切な関連語とコードとを含む検索式を提案する特許調査支援システムを開発した。本特許調査支援システムを用いれば、経験の浅い調査者でも、専門家並みの特

許調査が可能となる。

## 4.2 特許調査支援システムの機能

本システムは、キーワードを入力するだけで、キーワードとその関連語、およびコードを有する検索式を出力するシステムである。例えば、「OS タスク 速度」を入力すれば、検索式「(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)」を出力する。

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を有する。そして、検索式構築のための機能は、調査者の入力したキーワードから関連語を取得する関連語抽出機能と、キーワードとその関連語（以下、「検索語」という）から適切なコードを取得する特許コード抽出機能と、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で構成される。システム全体の概要を図2に示す。

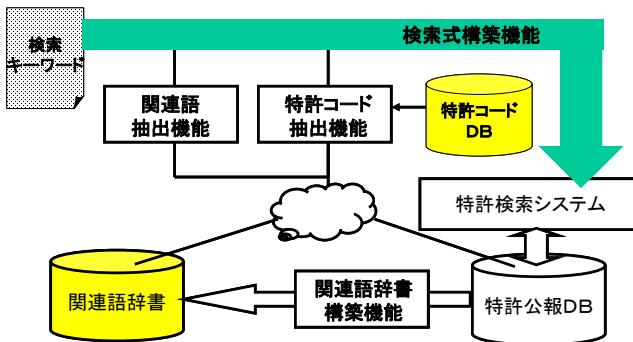


図2：特許調査支援システムの概要

### (1) 関連語辞書構築機能

本システムは、関連語辞書を構築する関連語辞書構築機能を有する。関連語辞書は、以下の二つの戦略のもと構築した。第一は特許公報の構造に着目する戦略であり、第二は特許公報に特有の手がかり表現に着目する戦略である。

第一の構造に着目する戦略において、例えば、請求項間の従属関係のうち、内的付加の従属関係にある従属請求項（例えば、「前記弾性体は、バネであることを特徴とする請求項1記載の〇〇装置」）の文から、構成要素名（例えば、「弾性体」）を上位語とし、構成要素の具体例（例えば、「バネ」）を下位語として抽出する。また、例えば、「【符号の説明】に記載されている構成要素を取得し、明細書の実施の形態における構成要素の具体例の例示文（例えば、「切断

手段は、例えば、カッター、レーザー、水等である。）」から、上位語（例えば「切断手段」）と下位語（例えば、「カッター」、「レーザー」、「水」）を抽出する。【符号の説明】に挙げられている構成要素は、特許公報において重要な用語であり、本手法は有効である。

第二の手がかり表現に着目する戦略において、例えば、明細書中の「HDDなどの記録媒体」の表現から、上位語「記録媒体」、下位語「HDD」を抽出する。ここで、「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。

### (2) 関連語抽出機能

本機能において、まず、入力されたキーワードの同義語を関連語辞書から取得する。次に、入力されたキーワードとその同義語の上位語、下位語を関連語辞書から取得する。そして、取得したすべての用語を関連語として出力する。

### (3) 特許コード抽出機能

本機能の動作する前に、調査者が、キーワードと関連語抽出機能が出力した関連語から、調査したい技術に関連する用語（検索語）を手動により選別する。そして、本機能において、特許公報ごとに特許公報内の全用語における検索語の出現割合（以下、「用語の出現率」という）を算出し、コードごとにそのコードが付与された特許公報の用語の出現率の和（以下、「コード関連度」という）を算出する。そして、コード関連度が高いほど適切なコードであるとした。コード関連度とは、用語とコードとの関連の度合いを示す情報である。

### (4) 検索式構築機能

本機能は、関連語抽出機能が出力した検索語と、特許コード抽出機能が出力した1以上のコードのうち、調査者が選択したコードとを用いて、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。なお、本システムにおいて、コードの選択を助けるために、コードの意味も出力される。

### 4.3 特許調査支援システムの評価結果

本システムの有用性について、以下のように評価した。本検証における特許検索は、特許電子図書館（IPDL）で行い、用語の検索対象を「要約+請求の範囲」とした。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。なお、本検証で調査対象として適切な公報（以下、「適合公報」という）は、検索結果である特許公報を検証者が読んで確認したものである。

まず、関連語抽出機能が出力する関連語の有効性について検証した。検証者が入力したキーワードのみで構成された検索式「OS\*タスク\*速度」で検索した結果、41件ヒットし、うち適合公報は13件であった。そして、キーワードと取得された関連語のうち、検証者が選択した用語からなる検索式「(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)」で検索した結果、476件ヒットし、うち適合公報は48件であった。

これより、本システムが提案する関連語が、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否かについて、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」という）が向上することが確認できた。つまり、本システムの提案する関連語を用いることで適合数が13件から48件となり、適合公報を多く取得できたことが分かる。

次に、本システムが提案する検索式の有用性について検証した。本検証では、IPCを用いて行った。本システムが出力した検索式「(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)」で検索した結果、39件ヒットし、うち適合公報は35件であった。このことより、本システムが提案する検索式を用いることで、キーワードのみの検索式の場合と比較して、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。以上、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。

## 5. 特許書類半自動生成システム

### 5.1 特許書類半自動生成システムの概要

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。明細書設計書は、権利化したい発明の内容を記載した文書であり、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」などが記載された文書である。「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されたとしても、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半自動生成できる。

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3参照）。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。

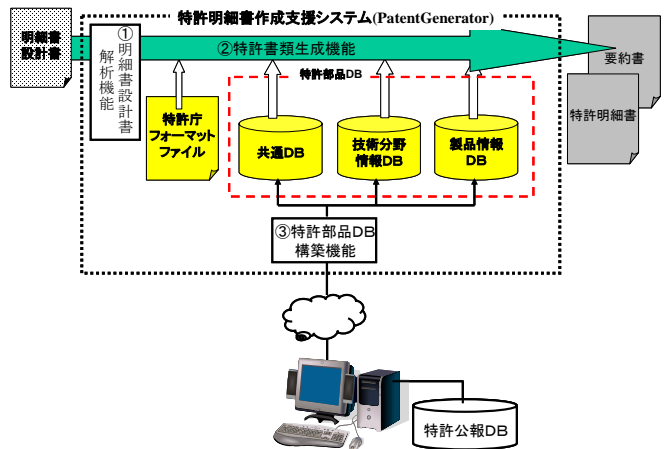


図3：PatentGeneratorの構成図

## 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、または説明手順の情報を格納している。

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、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を示す文等からなる。

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、製品の部品を説明する文や、発明を構成する構成要素の実現手段を説明する文等からなる。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（表 2 参照）。

表 2：製品情報 DB の例

名称	実現手段
格納	<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。
受付	入力手段は、テンキーやキーボードやマウスやメニュー画面によるもの等、何でも良い。<構成要素>は、テンキーやキーボード等の入力手段のデバイスドライバや、メニュー画面の制御ソフトウェア等で実現され得る。
受信	<構成要素>は、無線の通信手段が好適であるが、放送を受信する手段や有線の通信手段でも実現可能である。

## 5.3 特許書類生成機能

特許書類生成機能において、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、1 以上の構成要素名とそれらの階層関係、および技術用語を取得する。そして、構成要素名、技術用語を用いて、特許部品 DB 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。

詳細には、特許書類生成機能は、構成要素名の階層関係から、装置の構成を説明する文を生成し、出力する。装置の構成を説明する文とは、例えば、「特許価値算出装置 1 は、特許情報格納部 101、特許属性情報格納部 102、中間価値情報算出部 103、特許価値算出部 104、特許価値出力部 105、を具備する。」である。

また、特許書類生成機能は、明細書設計書を構成する各請求項を、構成要素ごとの説明文に分割し、実施の形態における表現に変更して、出力する。そして、構成要素ごとの説明文内の構成要素名、技術

用語をキーとして、特許部品 DB を検索し、構成要素の実現手段を示す文や、技術用語を詳細に説明する文などを取得する。そして、構成要素ごとの説明文の直後に、検索した文を出力していく。本システムの出力例を図 4 に示す。

(実施の形態 1)

本実施の形態において、分類されている特許に対して、分類毎の価値に基づいて、経済的重みを取得し、かつ、特許明細書、請求の範囲を解析し、法律的、技術的価値を算出する。そして、3 方向の総合的な特許価値を自動算出する。CL1-中間価値情報算出部 103 は、特許特性値を算出する、CL2-特許特性値の具体例特許価値算出装置 1 について説明する。また、本実施の形態において、<実施の形態 1 の前置部 なし>の特許価値算出装置 1 について説明する。

図～は、本実施の形態における特許価値算出装置 1 のブロック図である。

特許価値算出装置 1 は、特許情報格納部 101、特許属性情報格納部 102、中間価値情報算出部 103、特許価値算出部 104、特許価値出力部 105 を備える。中間価値情報算出部 103 は、特許書類解析手段 10301、特性値算出手段 10302、中間価値情報算出手段 10303 を備える。特性値算出手段 10302 は、発明本質抽出性値算出手段 1030201、発明展開性値算出手段 1030202 を備える。

特許情報格納部 101 は、1 以上の特許書類と特許書類の属性を示す特許属性を有する特許情報を格納している。

特許書類とは、特許請求の範囲、明細書、図面、要約書である。

特許属性とは、技術用語、IPC、Fタームなど、分類のための情報である。

格納しているとは、不揮発性の記録媒体でも良い。

特許情報格納部 101 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。

特許情報格納部 101 に〇〇情報が記憶される過程は問わない。例えば、記録媒体を介して〇〇情報が特許情報格納部 101 で記憶されるようになってよく、通信回線等を介して送信された〇〇情報が特許情報格納部 101 で記憶されるようになってよく、あるいは、入力デバイスを介して入力された〇〇情報が特許情報格納部 101 で記憶されるようになってよい。

図 4：PatentGenerator の出力例

## 5.4 特許部品 DB 構築機能

特許部品 DB 構築機能は、以下の動作をする。まず、特許公報の【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。」などの文（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。この文は、製品情報 DB のレコードとなる。

また、特許部品データベース構築機能は、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、共通情報 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。

### 5.5 特許書類半自動生成システムの評価結果

10 件の特許明細書作成を、PatentGenerator を用いて行った。その際の特許明細書の生成率は平均 48.1% であった。また、特許出願書類の作成効率も 1.8 倍に向上した。また、PatentGenerator により、特許明細書の品質が 1.4 倍に向上した。さらに、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チェック時間が、平均 1.5 時間から 1 時間に削減された。なお、特許明細書の品質の評価は、後述する特許書類品質評価システムを用いた。

## 6. 特許書類品質評価システム

### 6.1 特許書類品質評価システムの概要

特許書類品質評価システムは、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解析することで得られる約 100 の特許メトリクスから、特許書類の品質を評価する（図 5 参照）。

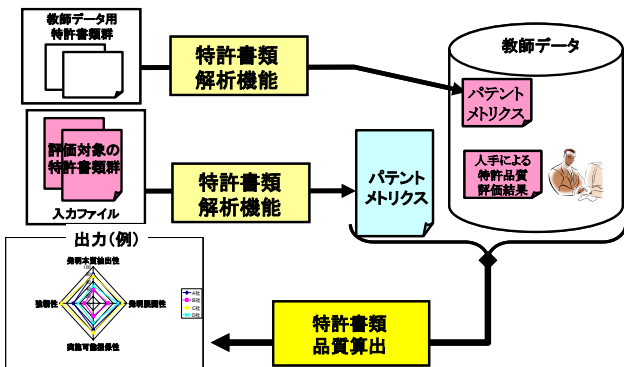


図 5：特許書類品質評価システムの全体構成

本システムにおいて、弁理士等の特許書類品質の評価者は、約 100 の特許メトリクスを考慮して、特許書類の品質を評価している、と仮定している。そして、この仮定の基、教師データベースに、特許書類から取得した特許メトリクスと、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 の特許メトリクスを抽出し、それらの特許メトリクスと教師データベースから、機械学習機能を用いて、評価対象の特許の価値を自動算出する。そして、本評価結果は、評価者の評価方法に従うこととなる。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) のアルゴリズムに基づく<sup>6)</sup>。

本システムにおける出力例は、図 6 である。特許書類品質評価システムにおいて出力できる評価値は、特許価値（全体の価値）、および特許請求項における限定度合いを示す「発明本質抽出度」、特許請求項の展開度合いを示す「発明展開度」、発明の詳細な説明における記述詳細度を示す「明細書開示度」、事前の特許調査を十分に実施したかどうかを示す「強靭度」の特許価値を示す指標である。

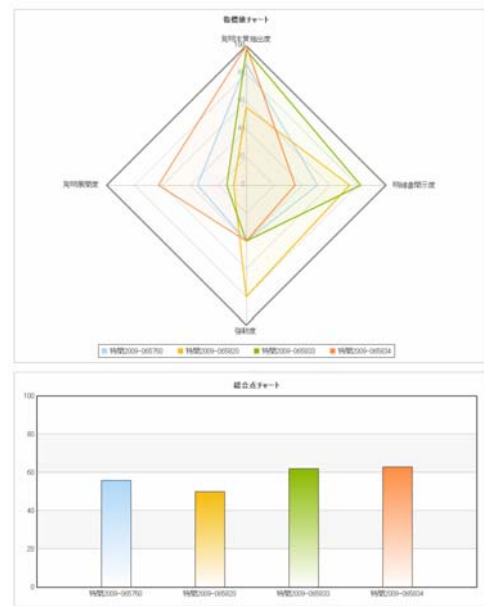


図 6：品質評価結果

### 6.2 パテントメトリクス

パテントメトリクスは、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である。特許書類品質評価システムでは、このパテントメトリクスを、評価対象となる特許書類から特許書類解析機能により取得する。

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。「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 特許書類品質評価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 点満点での評価である。本評価は、特許明細書品質評価である。評価結果を表 3 に示す。評価結果によれば、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。

表 3：本システムの評価結果

	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

### 7. おわりに

以上、統合的な特許工学ツールを実現し得る特許記述言語 (Patent Markup Language (PML)) について提案した。また、PML を用いて、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効果について説明した。

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。

### 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) 特許庁発行「公報仕様 特許、実用新案 第 4 版」
- 3) [http://www.patentisland.com/Patent\\_Strategy\\_Engineering/](http://www.patentisland.com/Patent_Strategy_Engineering/)
- 4) <http://www.patentcreator.biz/>
- 5) <http://www.intechstra.com/stravision/evaluation.htm> (StraVision)
- 6) <http://chasen.org/~taku/software/TinySVM/>