

特許調査業務を改善する特許読解支援システム—特許情報と技術者を近づけるための技術—

田中 一成^{†1} 池田 紀子^{†1}

^{†1}(株) 富士通研究所

特許は企業活動にとって欠かすことのできない財産であり、開発した技術の資産化と資金調達を含む経営判断においても重要である。特許を生み出したり、活用したりするためには特許調査を行う必要があり、その過程で特許文書を読むという作業は避けて通れない。一方で特許文書には独特の読みにくさがあり特許文書を読み解く作業が負担になってしまう。特に特許文書の独特の書き方に不慣れた技術者にとっては深刻な問題である。本稿では、特許文書を読むノウハウを機能として作り込むことによって、特許文書を読む作業を支援するシステムを構築したので報告する。さらに、化学分野に特化した特許読解支援システムの応用を通してさまざまな情報を統合し利活用する可能性について述べる。

1. はじめに

企業活動においては、技術内容を理解できる技術者(開発者、研究者)自身が特許文書を読む必要性がますます高まっている。それは、他社にない特許を出願したり、他社特許に抵触するリスクを回避したりするために、あるいは他社がすでに開発している技術を知らないまま類似の技術開発に労力を費やしてしまうことを避けるためである。

しかし、特許文書には独特の読みにくさがあるために、読んで理解する業務が本務を圧迫するほど大きな負担になるという問題が顕在化している。本稿では、特許文書に不慣れた多くのユーザの特許読解を支援するため、特許文書を読むために必要なノウハウを自然言語処理や機械学習などの人工知能技術の実践的活用により機能として実現した特許読解支援システムについて報告する。本システムを利用することで、現状の業務を効率化するとともに、特許調査への心理的負担を低減させるという効果もあった。

2. 特許調査の実際

2.1 特許調査の種類

特許調査は、目的によって、以下のように分類することができる。

• 公知例調査

特許を出願しようとする場合や審査請求をしようと

する場合に行うもので、すでに同様の技術が特許出願されていないかどうかを調べるものである。

この調査は、特許の発明者となる研究開発部門で行う場合もあれば、知財部門が行ったり、外部へ委託したりすることもある。

• 侵害回避調査

製品やサービスを事業化する前に行うもので、他社の特許を侵害していないかを調べるものである。

この調査は知財部門が行う以外に、開発現場の技術者が行う場合がある。

他社特許を知らずに侵害してしまうと、多額の損害賠償を請求されることになったり、製品の差し止めを受けたりすることもあるため企業にとってはリスクを回避するために重要な調査である。

• 動向調査

研究の企画段階などで、研究分野の出願動向を調査するもので、特許に付けられた分類や出願人(企業)といった定型情報を集計する比較的浅い分析から、特許の文書を人手で読み込んで分類するなどの深い分析までさまざまなものがある。

このほかにも、特許の権利維持の判断や、売買を行うときの価値評価のための調査などもある。

いずれにしても、最終的には見つけた特許1件1件を精査する必要が出てくるため、技術内容を理解できる技術者が特許を読むという作業は重要である。

2.2 特許文書の構造

特許文書は以下のような構造で書かれており、特許調査の目的によって読むべきところも異なる。

- 書誌事項（出願番号や特許の分類など）
- 要約（発明の概要）
- 請求項（どんなものが権利範囲に含まれるか）
- 産業上の利用分野（何についての発明か）
- 従来技術（関連の特許や論文には何があるか）
- 解決すべき課題（どんな課題を解決しているか）
- 解決手段（どのような方法で解決しているか）
- 実施例（実際にはどのようにして実施するか）
- 効果（発明の利点は何か）

たとえば、公知例調査では、解決手段や実施例が重要になるが、侵害回避調査では、請求項が問題になる。動向調査の場合は書誌事項の情報を使って集計を行ったり、要約に出現するキーワードを用いて分析を行ったりもする。

2.3 特許調査システムの特徴

特許調査では、調査の目的にもよるが、蓄積された膨大な特許情報から検索して関連のある特許の集合を抽出し、抽出した集合を精査したり、分析にかけたりする。

特許検索ではおおむねどの調査でも、検索条件を探すための予備検索、本検索、検索結果の精査や分析という過程を経る。特許文書には、IPCやFIなどの特許独自の分類コードが付けられている。知財部門や調査会社のような特許検索の専門家はキーワードだけでなく、こうした分類コードを予備検索で探し、本検索のための検索式を作成することで精度の高い検索を行う。予備検索でヒットするものを確認する段階や、本検索後に精査する段階で特許文書を読むことになる。動向調査では、本検索により抽出した集合を分析ツールにかけることによって、分析を行う。

特許検索の世界では、Webのようなランキング検索よりも、検索条件を定めて集合を抽出し、抽出した集合の内容をすべて精査するといった検索方法のほうが主流である。これは、調査した範囲を明確にする必要があるからである。侵害回避調査のように調査範囲を明確にする必要がある場合には、このような検索方法が主となるが、予備検索や公知例調査では、キーワードや文書を入力として類似する特許文書をランキングで表示する検索方法も実用化されている。本検索では検索漏れが問題になるが予備検索では多少の漏れは問題にならないためである。公知例調査では、特許検索に不慣れな技術者が、

厳しい検索条件を与えて読める件数まで絞り込むより、ランキング検索の上位を見るほうが、精度が高い場合もある。

分析は、定型情報を集計するものから、テキストマイニング技術を応用して、集合の概要や特許間の関係性を可視化する技術[1]も実用化されている。

2.4 特許調査システムの課題

特許調査には、特許知識の豊富な知財部門だけではなく、技術内容を理解できる技術者も参加しなければならない。

検索技術が進歩して、特許を検索する作業は技術者にも可能になってきたが、特許を読み解く作業は依然としてユーザ（技術者だけでなく知財部門、企画部門やSE部門なども）の力に頼らざるを得なかった。

特許文書には独特の読みにくさがあるために、検索によりヒットした特許文書を精査する業務が負担になっており、ユーザが難解な特許文書を読む負担を軽減する技術の開発が望まれていた。

**特許に不慣れな技術者にも
特許を読めるようにする！**

3. 特許読解支援システムの実現

3.1 読解支援の必要性

特許文書における独特の読みにくさはさまざまあるが、たとえば以下のようなものがある。

- 権利範囲を書く請求項では特有の言い回しで書かれる
- 請求項は、従属関係を意識しないで個々に読んでも完結しない
- 権利範囲を広くするために意図的に抽象的な言葉で書かれる
- 発明のポイント以外にも大量の文書が記述される

もちろん、社内教育などを通じて技術者が特許文書を読み解くレベルを上げることでこの問題に対処することも可能ではあるが、個々人にとっては年間に数回程度の特許調査をするためだけに教育に多くの時間をかけ、ノウハウを習得することは難しい。

特許特有の読みにくさに対するプラクティスから特許読解支援技術が生まれた。

3.2 特許読解支援システムの機能

特許文書に不慣れな多くのユーザーの特許読解を支援するために、特許文書を読むために必要なノウハウをいくつかの技術（請求項の構造解析技術、請求項の従属関係解析技術、具体説明情報抽出技術、特定観点情報抽出技術など）を組み合わせることにより機能として実現した特許文書の読解支援システムを構築した。

3.2.1 請求項の構造解析

請求項には、**図1**の例のように一文で権利範囲を書く慣習があるため、文書が複雑で難解なものになるという読みにくさがある。

特許文書を読むためのノウハウはいくつかある。たとえば、請求項を読む場合には、「□□□において、△△△と、○○○と、×××とを具備した☆☆☆装置。」といった典型的な表現パターンに沿って「と、」で改行を入れて構成要素ごとの記述を見やすくすることができるというものがある。請求項の書き方にはさまざまな表現のバリエーションがあるため、特許に慣れた専門家はバリエーションに合わせてうまく構造化することで請求項を読みやすくしている。

請求項に特有の言い回しをシステムに覚え込ませ、ユーザーが頭の中で読み解いていた思考をシステム側で行うようにして、**図2**のように解析結果を可視化する。これにより、発明の前提条件（**図2**では「おいて書き」と表記）、構成要素、構成要素の説明といった内容を素早く読み取ることができるようにした（色とアンダーラインについては後で説明する）。このような文書の構造化

特許文書を読むために必要な
ノウハウを人工知能技術
により機能として実現した！

技術により、請求項の言い回しを知らずに読もうとしていた技術者にとっても、ずっと内容が理解しやすくなった。

3.2.2 請求項の比較表示

特許の請求項は、出願して公開されてから特許庁で審査が行われる過程で、審査官と出願人のやり取りによって補正が行われる。そのため、請求項は、出願した時点と登録になった時点では内容が異なっている場合が多い。請求項では、特許の権利範囲に含まれる発明の構成要件が規定される。そのため、一般に請求項に限定が少ないほど（抽象的なほど）権利範囲が広いことになる。公開時と登録時の請求項の差は、審査官の拒絶理由に対応して権利範囲が限定された結果である可能性が高い。したがって、この差は、権利範囲を読み解くには重要な情報である。

そこで、請求項の構造化を応用して、公開公報と登録公報の請求項を自動で対応付け、差分を抽出して表示することで、権利化のポイントを明確化できるようにした。

図2の例について**図3**のように公開時の請求項を示すとともに**図4**のように、その特許の公開時の請求項と、登録時の請求項を対応付けて変化を見やすく整理して表示する。**図4**の例では、登録時に「表示手段」「関係抽出」「距

検索要求に適合する情報を検索する情報検索装置において、少なくともテキストを含む情報を記憶する情報記憶手段と、複数の単語を検索キーとして前記情報記憶手段から情報を検索する情報検索手段と、該情報検索手段で検索された情報から検索キーとして用いられた複数の単語間の関係を抽出する関係抽出手段と、前記情報検索手段による検索結果と前記関係抽出手段で抽出された関係とを対にして表示する表示手段を有し、前記関係抽出手段は、前記複数の単語が1文中に存在する場合は該1文を形態素解析して単語間の関係を抽出し、前記複数の単語が1文中に存在しない場合には、それぞれの単語を含む文と文の間の距離を算出することを特徴とする情報検索装置。

図1 請求項の例

【請求項1】（独立項）公開公報の請求項1と類似

おいて書き 検索要求に適合する情報を検索する情報検索装置において、

要素1 少なくともテキストを含む情報を記憶する情報記憶手段と、

要素2 複数の単語を検索キーとして前記情報記憶手段から情報を検索する情報検索手段と、

要素3 該情報検索手段で検索された情報から検索キーとして用いられた複数の単語間の関係を抽出する関係抽出手段と、

要素4 前記情報検索手段による検索結果と前記関係抽出手段で抽出された関係とを対にして表示する表示手段を有し、

要素の説明 前記関係抽出手段は、前記複数の単語が1文中に存在する場合は該1文を形態素解析して単語間の関係を抽出し、

要素の説明 前記複数の単語が1文中に存在しない場合には、それぞれの単語を含む文と文の間の距離を算出することを特徴とする情報検索装置。

図2 請求項（登録時）の構造解析例

公開公報の請求項 1	
おいて書き	検索要求に適合する情報を検索する情報検索装置において、
要素 1	少なくともテキストを含む情報を記憶する情報記憶手段と、
要素 2	複数の単語を検索キーとして前記情報記憶手段から情報を検索する情報検索手段と、
要素 3	該情報検索手段で検索された情報から検索キーとして用いられた複数の単語間の関係を抽出する関係抽出手段
を有することを特徴とする 情報検索装置。	

図 3 請求項（公開時）の構造解析例

公開公報の請求項 1 と登録公報の請求項 1 の対比

構成要素	情報記憶手段	情報検索手段	関係抽出手段	表示手段	関係抽出	距離算出
公開公報の請求項 1	少なくともテキストを含む情報を記憶する情報記憶手段	複数の単語を検索キーとして前記情報記憶手段から情報を検索する情報検索手段	該情報検索手段で検索された情報から検索キーとして用いられた複数の単語間の関係を抽出する関係抽出手段	---	---	---
登録公報の請求項 1	同一	同一	同一	前記情報検索手段による検索結果と前記関係抽出手段で抽出された関係と対して表示する表示手段	前記関係手段は、前記複数の単語が 1 文中に存在する場合は該 1 文を形態素解析して単語間の関係を抽出し、	前記複数の単語が 1 文中に存在しない場合には、それぞれの単語を含む文と文の間の距離を算出する

図 4 請求項の比較表示例

離算出」について限定が増えており、権利化の争点になったと推測できる。

3.2.3 請求項の従属関係解析

請求項には従属関係があり、その関係を意識しないで個々の請求項を読んでも理解することが難しい。

請求項の従属項では、「～ことを特徴とする請求項 1 または 2 に記載の情報検索装置。」のように言語表現でさらに発明の特徴が付け加えられるので、このような記述パターンを蓄積して従属関係を抽出できるようにした。抽出は、正規表現の 7 種類のパターンを用いることによりほとんどの場合をカバーすることができた。たとえば、「/請求項 ([0-9 0-9]+) に? (記載)? [のした]+/」や「/請求項 ((請求項)? [0-9 0-9]+ [, , 又は またおよび及]+) (請求項)? ([0-9 0-9]+) /」といったようなパターンである。

請求項の従属関係を記述している表現を抽出し、図 5 のように従属関係を構造化して表示する。これにより、独立項から順に読み進めることができるようにした。従属関係は、1 対多であり、この例の請求項 3 では「請求項 1 または 2 に記載の情報検索装置。」と書かれている。請求項 1 にも 2 にも従属しているため、図 5 のように木構造に展開すると、2 回出現することになる。

3.2.4 具体説明情報抽出

請求項では、権利範囲を広くするために意図的に抽象的な言葉を使って書くことが多い。

請求項で書かれている抽象的な言葉は、具体的な意味

が特許文書の中の実施例に書かれる場合が多い。そこで、情報抽出技術を応用して抽象的な言葉の具体的な意味をあらかじめ抽出しておいて、いつでも参照できるようにした。

たとえば、請求項では「複数の単語間の関係」というように書かれていても、実施例では『検索結果から単語列と抽出規則との比較によって関係を抽出し、「ヲ格」「ガ格」などを関係として用いたが、この関係としてこれ以外にも文章を構文・意味解析して得られる格関係などを用いても良い。』と説明されていたりする。

「として」や「とは」といった手掛かり語と記述のパターンを利用して、抽象的な言葉とその説明を抽出しておき、特許文書中で出てくる抽象的な言葉にリンクを張る。たとえば図 2 では、青字でアンダーラインの語句（「検索要求」「検索装置」「検索キー」など）にリンクを張り、クリックすることで具体的な説明を参照することができるようにした。この機能により、わざわざ実施例の中を探さ

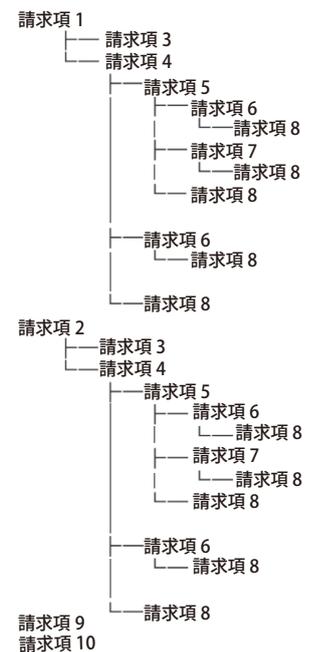


図 5 従属関係の表示例

なくてもユーザは抽象的な言葉に惑わされずに読み進めることができるようにした。

具体説明情報の抽出では、手掛かり語の直前を抽象的な語として抽出するが、その具体的な説明がどの部分かまでは特定が難しく、現状では手掛かり語を含む1文すべてを抽出するようにしている。構文解析や、言葉を意味に踏み込んで扱うことができるオントロジー（概念体系）を応用することにより、さらに簡潔な情報提供が可能になるものと考えている。

3.2.5 特定観点情報のハイライト

筆者らは、人工知能技術を応用し、特許に書かれている文書を係り受け解析し、自然言語の表現知識を利用して、その特許が何についての技術か（発明の対象）、どういった課題を解決しようとしているのか（発明の課題）といった情報を抽出する技術を開発してきた[2],[3]。特許文書の技術分野や観点によるが、8割前後の精度で抽出できるようになっている。この技術を応用し、特許文書中で発明の対象や課題を記述している部分を自動的にハイライト表示する。これにより、ユーザは読むべきところに当たりを付けることができ、目的の情報へ効率良くたどり着けるようになった。

特定観点情報抽出のさらなる精度向上には意味に踏み込んだ解析が必要であると考えている。

3.2.6 その他の機能

このほかにも、以下のような機能を実装しユーザの負荷軽減に貢献している。

- 各請求項から発明の構成要素名を抽出して表形式にまとめて表示することで、特許の全体像を把握し、詳細に読むべき内容を把握できるようにした。
- 特許文書全体を構造化して、請求項、従来技術、効果などの中から必要な部分だけを表示することで、読んでいる場所を分かりやすくした。
- 請求項の構成要素の中で最も特徴的な構成をハイライトする（図2においてピンク色で表示されている「抽出」のようなもの）。

4. 特許読解支援システムの実用

4.1 特許調査システムの構成

読解支援技術を実業務の中で使って評価してもらうために、検索システムとの連携が不可欠であった。

そこで、当社で研究開発しているテキストマイニング技術を応用した特許検索システムを構築し、その公報表示機能として特許読解支援システムを位置付け、検索か

ら読解までをシームレスに連携したシステムを構築した。図6にシステムの構成を示す。

このシステムでは、単純なキーワード検索のようなシンプルなGUIにすることによって技術者が馴染みやすいようにした。これにより、知財部門でも研究開発部門でも読解支援システムを実業務の中で使ってもらえるようになった。

実業務に耐えうるシステムにするために、特許の公報データおよび特許の審査経過のデータを月に1度の頻度で更新する体制を整えた。

また、ユーザからのフィードバックを受けながら実験的にさまざまなデータや機能を取り入れた。機械学習技術を応用した特許評価機能もその1つであり、特許の引用関係や出願人が行うアクション（審査請求や補正など）といった情報を特徴として、人手で行う特許の価値評価を自動化する機能を実用化した。

4.2 実用性の評価手法

特許読解支援システムの実業務への貢献度合いを定量化するため、ユーザへのアンケート調査とアクセスログの解析を行った。

アンケート調査は、読解支援のトップページにリンクを貼り、知財部門11名、研究開発部門14名、その他6名の合計31名から回答を得た。

4.3 評価結果

上記のようなシステムを開発して社内業務で利用できるようにすることで、社内の研究開発部門、知財部門のほか、企画部門やSE部門などにも広く活用されている。

アンケート調査を実施した結果では、平均で特許を読

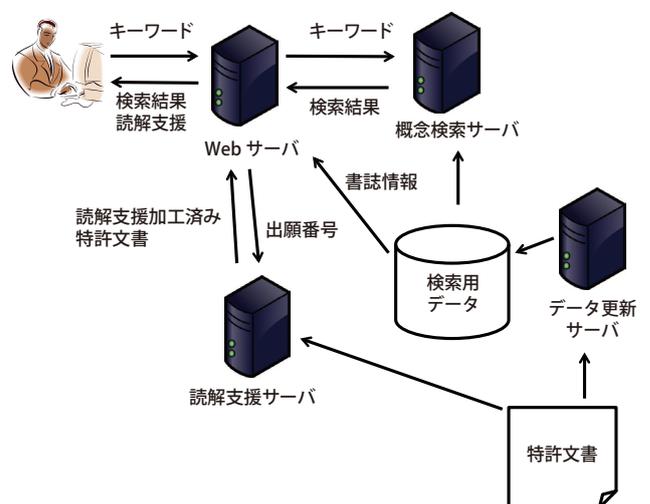


図6 特許調査システムの構成図

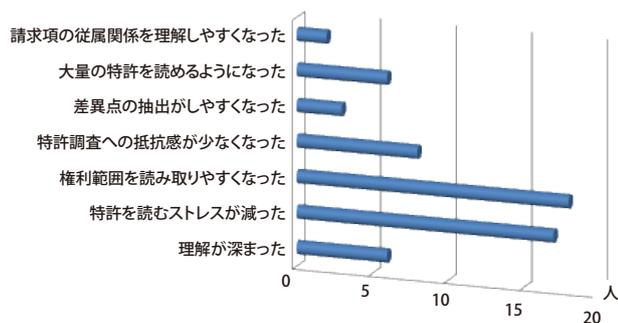


図7 読解支援の評価

むのにかかる時間が約40%程度削減されることが分かった。

特許読解支援システムは、試作されたのち社内業務で活用されている。現場からは特許文書の解析誤りについてのフィードバックを多く受けている。現状のシステムでは、その都度解析ルールを修正するしかないが、手作業によるルールの修正には限界がある。ユーザの操作によって、解析結果が正解か不正解かという情報を蓄積し、機械学習によってルールを自動的に再構築する仕組みが必要だと考えている。

アンケート調査では、効率化以外にも図7に示すように、半数以上のユーザが以下のような効果を感じていると回答している。

- 権利範囲を読み取りやすくなった
- 特許を読むストレスが減った

このほかにも、「特許調査への抵抗感が少なくなった」「大量の特許を読めるようになった」といった回答も多く、特許調査への心理的な負担を低減させる効果が確認された。また、特許文書を印刷する枚数も約半数の人が2割以上少なくなったと回答した。

2年間の読解支援システムへのアクセス回数は15万回程度で、実業務で使うツールとして定着した。

社内利用で効果が実証できた特許読解支援システムは、2012年度に人工知能学会の現場イノベーション賞(銀賞)をいただいた。ビジネスの現場における実問題に関して、人工知能技術により解決したことが認められたためである。

5. 特許読解支援技術の発展

これまでは一般的にどの分野の特許文書にも利用できる技術を開発し、読解支援システムに実装してきた。

しかし、それだけで不十分な場合は、読解の妨げが生じていた。たとえば、化学分野の特許文書では、請求項の書き方がほかの分野とは異なることや、化合物(化学

物質)名の羅列が多いことが原因である。

そこで、これまでの特許読解支援システムが特に苦手としていた化学分野に特化した技術開発を行っている[4]。

化学分野特許の 読解も支援する！

5.1 化合物名の認識

化学分野に特化した読解支援を実現するためには、化合物名を認識することが重要となる。化学分野の特許文書を一般的な形態素解析にそのまま通してしまうと、化合物名の途中で分断されて抽出されてしまうことが多い。

しかし、化合物名の認識を化合物辞書だけでカバーするには以下のような問題が指摘されている[4],[5]。

- 日々、新物質が誕生するため、辞書の作成には労力が必要
- 表記に関する基準や方針が時代とともに変化
- 医薬品などの商品名が増加
- 表記法に自由度があるために、書き手が勝手に作成
- 特許文書中で誤名称が発生

化合物の構造は置換基(原子の集まり)の組合せによって構成されている。化合物名の命名は多様だが、置換基名を組み合わせる場合が多い(体系名)。そこで、置換基名に使われる文字列を特徴素として機械学習を用いることにより、化合物名を認識する方式について検討した。

化合物名の認識は、まず、化合物名を構成する文字のパターンにより候補を抽出し、これを機械学習(ロジスティック回帰)によって化合物名か否かを判別する分類器を作成した。4,000件の学習用データ(正例1,267,負例2,733)を用いて、分類器を作成したところ80.4%程度の精度で化合物名を認識できるようになった。化合物を認識できなかったものとしては、「アルデヒド」、「アルミナ」、「アセトン」など置換基の組合せから命名されていないものであった。また、同じものに複数の呼び方があったり、「-」などの記号の使い方によって起こる表記ゆれが存在したりする。そのため精度向上には、この表記ゆれをクレンジングする機能が必要である。

化合物名を認識できるようになったため、化合物名に一連のIDを付けて判別しやすくしたり、化合物名にさらに詳しい情報へのリンクを付けたりするようにした。

5.2 化合物情報の知識化

化学分野特許を読み解くには、化合物の構造や特徴を集約して可視化することが必要不可欠である。

有機化合物の情報を蓄積しているデータベースには、国内最大規模の化合物数（約350万件）を収録している日本化学物質辞書（日化辞）[6]がある。J-GLOBAL[7]サービスでは日化辞を統合し、有機化合物について、別称、分子量、構造情報、各種法規制などの情報を利用できる。

しかし、5.1節で述べたように新規物質の増加に化合物辞書が追いつかないという問題がある。

そこで、増え続ける特許文書から新規化合物情報を抽出する方式について検討を行った。化合物構造の網羅性を高めるため、特許文書から抽出される化合物名と化学式の組を利用した。さらに、化合物の命名規則を利用して、部分構造に分解し、各部分構造名とその化学式の組を利用した。

ポリエチレングリコール		
機能・用途	出現回数	代替物質の候補
可塑剤	687	ジブチルフタレート ジオクチルフタレート
界面活性剤	541	ポリビニルアルコール
結合剤	192	ヒドロキシプロピルセルロース ポリビニルアルコール カルボキシメチルセルロース

図8 化合物の機能・用途と代替の例

また、3.2.4節で述べた具体説明情報抽出技術を応用して情報を抽出することで、化合物の機能・用途と代替物質の候補も抽出できる。図8は特許文書から抽出した化合物「ポリエチレングリコール」について、機能・用途、機能・用途ごとの出現回数、代替物質の候補を示す。

5.3 複数の化合物情報の知識化

化学分野の特許では、図9のように化合物名が羅列されるという特徴がある。

このような場合、複数の化合物を認識するために共通構造とバリエーションを明確にすることにした。

そこで、5.2節と同様に部分構造に分解し、複数の化合物群に共通する構造をそろえて、表形式に整理する技術を開発した。図10は図9の例について、化合物名と部分構造名を抽出し、表形式に整理して表示した結果である。この例では、大きく2パターンに分類することができ、1つ例外的なものが混ざっていることも読み取ることができた。

5.4 オープンデータの活用

特許文書からの知識化では、部分構造と機能・用途などの情報がある。一方で、特定の化合物の構造情報や、数多く存在する別称を特許文書から正確に抽出することは難しい。日化辞には、化合物の構造と、別称の体系名や慣用名も収録されている。化合物辞書は日化辞やWikipediaがオープンデータ（機械判読に適したデータ

【0074】

中でも、1,1-ビス(4-ヒドロキシフェニル)シクロペンタン、1,1-ビス(3-メチル-4-ヒドロキシフェニル)シクロペンタン、1,1-ビス(4-ヒドロキシフェニル)シクロヘキサン、1,1-ビス(4-ヒドロキシフェニル)-3,3,5-トリメチルシクロヘキサン、2,2-ビス(4-ヒドロキシフェニル)アダマンタン、2,2-ビス(3-メチル-4-ヒドロキシフェニル)アダマンタン、1,3-ビス(4-ヒドロキシフェニル)アダマンタン、1,3-ビス(3-メチル-4-ヒドロキシフェニル)アダマンタン、1,1-ビス(3-メチル-4-ヒドロキシフェニル)シクロヘキサン、1,1-ビス(4-ヒドロキシフェニル)シクロドデカン、1,1-ビス(3-メチル-4-ヒドロキシフェニル)シクロドデカンが溶解性に優れるPC共重合体を与えるという点で好ましい。

図9 化合物名を羅列している例

化合物名	母核	結合位置	置換基数	母核炭素数	第1の置換基		第2の置換基		第3の置換基	第4の置換基	第5の置換基		
					母核	置換基1	置換基2	母核	置換基1	置換基2	母核	母核	母核
1,3-ビス(3-メチル-4-ヒドロキシフェニル)アダマンタン	アダマンタン	1-, -3-,	2	10	フェニル	3-メチル	4-ヒドロキシ	フェニル	3-メチル	4-ヒドロキシ			
2,2-ビス(3-メチル-4-ヒドロキシフェニル)アダマンタン	アダマンタン	2-, -2-,	2	10	フェニル	3-メチル	4-ヒドロキシ	フェニル	3-メチル	4-ヒドロキシ			
1,1-ビス(3-メチル-4-ヒドロキシフェニル)シクロドデカン	シクロドデカン	1-, -1-,	2	12	フェニル	3-メチル	4-ヒドロキシ	フェニル	3-メチル	4-ヒドロキシ			
1,1-ビス(3-メチル-4-ヒドロキシフェニル)シクロヘキサン	シクロヘキサン	1-, -1-,	2	6	フェニル	3-メチル	4-ヒドロキシ	フェニル	3-メチル	4-ヒドロキシ			
1,1-ビス(3-メチル-4-ヒドロキシフェニル)シクロペンタン	シクロペンタン	1-, -1-,	2	5	フェニル	3-メチル	4-ヒドロキシ	フェニル	3-メチル	4-ヒドロキシ			
1,3-ビス(4-ヒドロキシフェニル)アダマンタン	アダマンタン	1-, -3-,	2	10	フェニル	4-ヒドロキシ		フェニル	4-ヒドロキシ				
2,2-ビス(4-ヒドロキシフェニル)アダマンタン	アダマンタン	2-, -2-,	2	10	フェニル	4-ヒドロキシ		フェニル	4-ヒドロキシ				
1,1-ビス(4-ヒドロキシフェニル)シクロドデカン	シクロドデカン	1-, -1-,	2	12	フェニル	4-ヒドロキシ		フェニル	4-ヒドロキシ				
1,1-ビス(4-ヒドロキシフェニル)シクロヘキサン	シクロヘキサン	1-, -1-,	2	6	フェニル	4-ヒドロキシ		フェニル	4-ヒドロキシ				
1,1-ビス(4-ヒドロキシフェニル)シクロペンタン	シクロペンタン	1-, -1-,	2	5	フェニル	4-ヒドロキシ		フェニル	4-ヒドロキシ				
1,1-ビス(4-ヒドロキシフェニル)-3,3,5-トリメチルシクロヘキサン	シクロヘキサン	1-, -1-, -3-, -3-, -5-,	5	6	フェニル	4-ヒドロキシ		フェニル	4-ヒドロキシ		-3-メチル	-3-メチル	-5-メチル

図10 化合物群の比較例

形式で、2次利用が可能な利用ルールで公開されたデータ)として利用できる。

特許文書から抽出した情報とオープンデータを相補的に使うことにより、より網羅的で正確な知識を活用することが可能になると考えている。

6. おわりに

特許庁が発行した『戦略的な知的財産管理に向けて—技術経営力を高めるために—』という報告書[8]によると、「特許情報の活用に関する将来像(理想像)」として、多くの企業で「発明者自らが先行技術調査・分析を行い、特許情報を活用できるようにする」と回答している。企業にとっては、他社がすでに開発している技術へ無駄に投資をしてしまうのを回避するばかりではなく、技術者の一人ひとりが特許文書に親しむことで、特許になる技術とはどういうものか、自分の持っている技術をどのように主張すれば権利化し保護できるのかといった特許的な感覚を養うことができ、会社全体の知財力を強化することができる。

特許読解支援システムを利用することで、技術者が特許情報を読めるようになり、特許制度の目的の1つである、他社の研究開発成果を利用して産業を発達させるための特許情報の活用がより一層促進される。特許情報として発行される研究開発成果をベースとして新しい技術の開発につなげることが可能となり、日本産業の発展にも寄与するものと考えられる。

今後は、人工知能関連技術を仲介として、特許情報とさまざまなオープンデータ、あるいは社内で持っているデータなど、多種多様なデータをつなげることで、単独のデータからは得られない有用な知識が得られるものと期待され、情報の有効活用の推進に寄与できるものと考えられる。本研究で、化学情報を扱ううちに、オントロジーのような意味構造の必要性が感じられるようになった。たとえば、上位下位概念や、置換基間の結合の

階層構造などである。すでに医療分野などでは、オントロジーのLinked Open Data化の研究が進められている[9]。特許情報を使ったオントロジー構築の研究も行われている[10]。このような研究成果も応用することで、さまざまな分野に適応した特許情報およびオープンデータなどの情報活用が可能になるものと考えている。

参考文献

- 1) 安藤俊幸: テキストマイニングを用いた効率的な特許調査方法, Japio YEAR BOOK 2015, pp.230-237 (2015).
- 2) 田中一成: 特許文書の多観点分類について, 情報処理学会研究報告自然言語処理 (NL) (47 (2004-NL-161)), pp.69-74 (2004).
- 3) 田中一成: 特許文書の多観点分類について, 情報処理学会研究報告自然言語処理 (NL) (4 (2008-NL-183)), pp.9-14 (2008).
- 4) 池田紀子, 田中一成: 特許文書からの化学物質情報の抽出, Japio YEAR BOOK 2015, pp.274-281 (2015).
- 5) 藤井 敦, 田中み子: 特許検索における化学物質名の異表記同定に向けた考察, Japio YEAR BOOK 2010, pp.182-187 (2010).
- 6) 日化辞, <http://dbarchive.biosciencedbc.jp/jp/nikkaji/desc.html> (2016年7月5日現在)
- 7) J-GLOBAL, <http://jglobal.jst.go.jp/> (2016年7月5日現在)
- 8) 特許庁: 戦略的な知的財産管理に向けて—技術経営力を高めるために— (2007).
- 9) 古崎晃司, 山縣友紀, 国府裕子, 今井 健, 大江和彦, 溝口理一郎: 医療知識基盤の構築に向けた疾患オントロジーのLinked Open Data化, 人工知能学会論文誌, 29巻4号, pp.396-405 (2014).
- 10) 難波英嗣, 乾 孝司, 岩山 真, 櫻井 孝, 橋田浩一, 藤井 敦: 特許分類コード体系に基づくオントロジーの構築—情報分野におけるケーススタディー, 言語処理学会第20回年次大会 (NLP2014). B2-5 (2014).

田中 一成 (正会員) tanaka.kazunari@jp.fujitsu.com
2001年北海道大学大学院工学研究科修士課程修了。同年(株)富士通研究所入社。テキストマイニング技術の研究に従事。その後、テキストマイニング技術を活かして特許の読解支援システムを開発。現在はLODの研究に従事。

池田 紀子 (正会員) nona@jp.fujitsu.com
1979年横浜国立大学教育学部化学科卒業。同年(株)富士通研究所入社。電子デバイス材料、並列処理、分子構造の研究に従事。高分子学会、人工知能学会、日本免疫学会各会員。技術士(応用理学/総合技術監理部門)。

採録決定: 2016年7月5日

編集担当: 浦本直彦 (日本アイ・ビー・エム(株))