

# 日本語概念辞書による特許作成支援とイノベーション支援

独立行政法人情報通信研究機構  
知識創成コミュニケーション研究センター  
MASTARプロジェクト言語基盤グループ  
風間 淳一

**PROFILE** 独立行政法人 情報通信研究機構 知識創成コミュニケーション研究センター MASTARプロジェクト言語基盤グループ 研究員。北陸先端科学技術大学院大学情報科学研究科助教を経て2008年8月より現職。自然言語処理の研究に従事。

✉ kazama@nict.go.jp

独立行政法人情報通信研究機構  
知識創成コミュニケーション研究センター  
MASTARプロジェクト言語基盤グループ  
村田 真樹

**PROFILE** 独立行政法人 情報通信研究機構 知識創成コミュニケーション研究センター MASTARプロジェクト言語基盤グループ 主任研究員。1997年京都大学にて日本学術振興会リサーチ・アソシエイト、1998年郵政省通信総合研究所入所を経て2008年より現職。自然言語処理の研究に従事。

✉ murata@nict.go.jp

独立行政法人情報通信研究機構  
知識創成コミュニケーション研究センター  
MASTARプロジェクト言語基盤グループ  
鳥澤 健太郎

**PROFILE** 独立行政法人 情報通信研究機構 知識創成コミュニケーション研究センター MASTARプロジェクト言語基盤グループ グループリーダー。1995年東京大学助手、北陸先端科学技術大学院大学准教授を経て、2008年より現職、自然言語処理の研究に従事。

✉ torisawa@nict.go.jp, <http://kccc.nict.go.jp/>

## 1 はじめに

特許は、産業技術の健全な発展や継承にとってなくてはならないものであり、特許情報処理の高度化の必要性はますます高まっている。本稿では、現在我々が開発している日本語概念辞書の特許情報処理への応用について、その展望を述べる。

## 2 キーワード想起支援と日本語概念辞書

現在、多くの人々が、様々なトピックに関する問題回避、あるいは、行動に関する方法やTipsについて情報を得るために、Web検索を利用している。しかし、そのような情報を得るためには、多くの場合、「意外」な

キーワードを入力する必要がある。例えば、「あじさい」には実は毒があり、実際に中毒事件も起きている。当然、「あじさい」と「毒」というキーワードを与えれば、通常のWeb検索でもその事実が見つかるが、重要なことは、こうした意外なキーワードはユーザの意識にない以上、システム側から提示する必要があることである。

我々は、こうした実生活でのキーワードの想起などを支援するため、Webから「生きた」シソーラスとその中の単語の間の意味的な関係や類似度を抽出・整理して「日本語概念辞書」を構築することを行っている。この日本語概念辞書には、通常のシソーラスに加えて、価値ある情報を効率良く提示するのに有用と考えられるいくつかの意味的な関係が定義されており、それぞれの語の関連語がそれにしなげって分類されている。現時点では、モノを利用したり対処したりする行為に関する情報が検索ニーズの大きな部分を占めているという仮定から、

「利用」「準備」「トラブル」「方法」「ツール」という5つの意味的關係を設定している。「利用」は、利用する行為自体（例えば、「あじさい」ならば「鑑賞する」）を表し、「準備」は、「利用」を行うためにその準備として必要な行為（例えば、あじさいの「栽培」）を表す。そして、これらの利用や準備といった行為を阻害する要因としての「トラブル」、それらの行為を行う際のTipsとしての具体的「方法」、有用な「ツール（材料）」というカテゴリがある。例えば、「毒」は、「あじさい」を、料理の見栄えを良くするため添える（鑑賞）際の「トラブル」であり、また、「レーザー手術」は「花粉症」の治療のための「方法」である。図1は、現在我々がこの日本語概念辞書を用いて開発しているキーワード想起支援システム<sup>1)</sup>の動作画面の例で、ユーザが入力したトピッ

ク「ダイエット」に利用できるツールや材料が、「トマト」のような意外なものも含めて提示されている様子を表したものである（「器具」などの語や吹き出し、語を囲む線はこの記事のために後から加えた）。意味的に類似した関連語は、単語間の類似度を利用することで、まとまって表示されるようになっており、欲しい関連語を容易に探すことができる。

また、各単語をクリックすると、その関連語も含めたWeb検索を行うほか、その語に類似した語を表示する機能もある。

また、この日本語概念辞書には、先に述べたとおり、シソーラスが含まれている。これは、各語に対して、その上位概念を表す語（上位語）（例えば、「東京大学」に

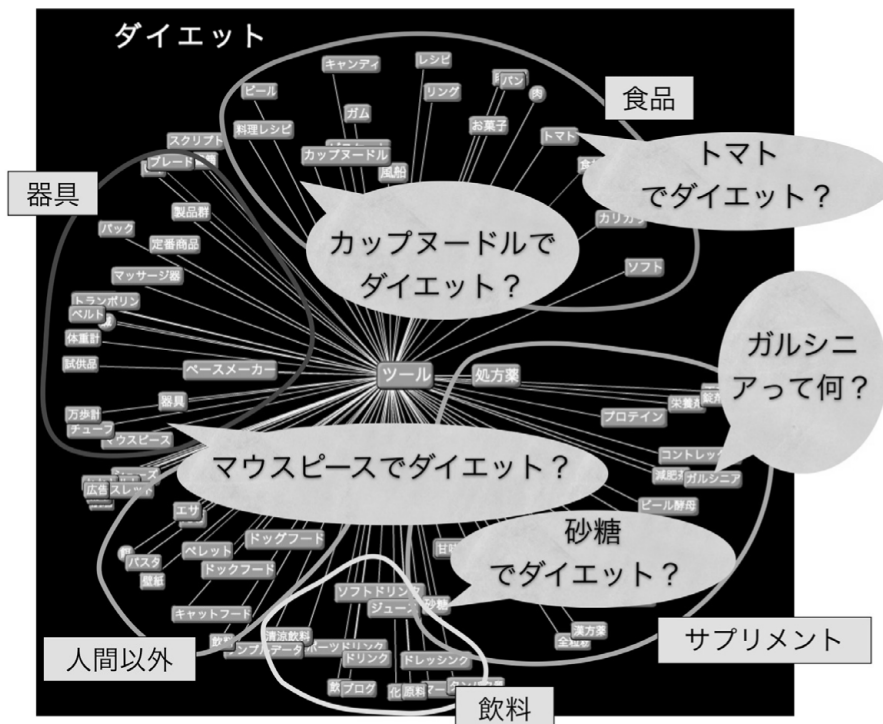


図1 キーワード支援システムの動作画面

<sup>1)</sup> 鳥澤健太郎、隅田飛鳥、野口大輔、風間淳一。自動生成された検索ディレクトリ「鳥式」の現状。言語処理学会第14回年次大会論文集，pp.729-730，2008。



対する「大学」を自動的に獲得することで作成している。上の想起支援システムでは、このシソーラスを利用することで、ユーザが入力したトピックが詳細すぎて十分に関連語が得られない場合にも対処できるようになっている。今年話題になった「農薬ぎょうざ事件」を例にとると、我々が開発に用いている昨年のWeb文書からは「ぎょうざ」のトラブルとして「農薬」を得ることはできなかった。しかし、「ぎょうざ」の上位概念に「冷凍食品」があることは得られており、「冷凍食品」のトラブルとして「残留農薬」があることも得られている。これらを組み合わせれば、「残留農薬」を「ぎょうざ」のトラブルとして、あたかも予測したかのように、提示することができる。

以上説明した日本語概念辞書は、大量のWeb文書やWikipediaの記事に自然言語処理を適用することで自動構築されており、現在128万語という大量のトピックを網羅している。我々は、この開発を通じて、この概念辞書やその獲得技術が、特許情報処理においても大いに役に立つであろうという感触を得ている。次の節では、我々がもっているいくつかのアイデアを述べたい。

### 3 特許情報処理への展開

特許は、これまでにない新規なアイデアによって成立する。これと、我々の元々の目的であった「意外なキーワードの提示」には何らかの関係性がありそうである。だとすると、開発した日本語概念辞書を特許情報処理に有効に利用することもできるかもしれない。また、ユーザの求める情報に注目して設定した「トラブル」の意味

関係は、まさに、ユーザのニーズ、「必要は発明の母」という言葉の「必要」、を考えるのに重要な概念であり、特許の発想を支援する際にも役に立ちそうである。以下では、検索と発想支援という二つの場面を考え、日本語概念辞書の利用方法を考える。

#### 3.1 類似特許検索

特許の申請のための書類の作成や、審査の際には、あるアイデアが既にある特許と重複していないかを調べるのが重要になる。そのときに必要不可欠なツールが検索システムである。文書検索で一般的に大きな問題になるのが、表記のゆれなど、同じ概念が違う言葉で表現されるという問題である。特許文書でも、用いる言葉を統制するようなことは行われていないため、検索の際に大きな問題になる。例えば、「ネット販売」で「オンラインショッピング」の特許を検索できないといった問題である。日本語概念辞書にある単語間の類似度を用いると、「ネット販売」に非常に近い語は「オンラインショッピング」であるといったことが計算できる。これを利用すれば、上のような表記ゆれなどに対応できると考えられる。さらに、単語類似度の基になっている確率モデルでは動詞の類似度も計算できるため、「水泳を習う」と「競泳を教わる」といった文間の類似度も計算することができ、これを応用すれば、文レベルでの表記ゆれの吸収も可能になると考えられる。

また、特許では、あるアイデアが既にある特許を包含しているか、並列しているか、あるいは包含されているかといった関係のチェックも重要になる。例えば、「梨」に対する新しいアイデアを思いついたとして、「豊水」(下位概念)に対する特許や、「リンゴ」(同位概念)に

対する特許、「バラ科植物」（上位概念）に対する特許が既にないかをチェックする必要がある。もしある場合には、そのアイデアは新規性がない、あるいは進歩性がないと見なされる可能性が大きいからである。

このようなチェックは非常に重要であるが、現状の検索システムではこのような関係を扱うようにはなっていないため、特許の申請や審査で、現在これが厳密に行われている訳ではない。真に新規で進歩的な特許のみを認めるためにはこのような検索を実現する必要があるといえる。日本語概念辞書の上位-下位関係を用いれば、このような、上位・同位・下位概念を判定することができ、また、同位概念は、単語の類似度を用いてもある程度判定することができる。

以上のように、概念辞書のような言語資源をうまく用いて検索システムを設計することで、より効率的で公平な特許システムを実現することができ、ひいては、社会の進歩にも大きく貢献できると考えている。

### 3.2 イノベーション支援

前節では、アイデアが既にある場合に、特に特許検索において、日本語概念辞書がどのように役立つかという展望を述べた。しかし、我々は、概念辞書や関連する技術は、アイデアそのものを生み出すときにも大いに役立つのではないかと考えている。以下で、いくつかのアイデアを述べる。

[ニーズ発見支援] 「必要は発明の母」というように、ある発明は、ある必要性がきっかけとなって生まれることが多い。また、技術の有用性は、それにいかに大きなニーズがあるかにかかっている。「トラブル」は、あるト

ピックに関する人々の不満や、落とし穴を表しており、それを解決する技術は有用であり特許となる可能性も高い。例えば、前述のキーワード想起支援システムで「飛行機」を検索すると、トラブルに「睡眠不足」や「騒音」等が出てくる。ここから、「席をカプセルホテルのようなベッド型のカプセルとし、さらにカプセルにノイズキャンセリング機能を持たせることで睡眠不足を解決する新しい飛行機」といったアイデアを思いつくかもしれない。このように、日本語概念辞書の「トラブル」を出発点にすることで、有用性の高い技術を効率よく得られる可能性がある。

[アイデア発想支援] 解決すべき問題が見つかったとしても、実際にうまくいく解決策を見つけることは非常に難しい。開発の途中では、当初考えていた方法ではうまくいかず、袋小路に陥ってしまうこともある。そのような場合、自分が愛着をもつ方法をいったん忘れて、可能性を一つずつ地道につぶしていくと打開策が得られることもある。日本語概念辞書にある単語類似度を用いて、類似した手法を列挙したり、まだ試していない可能性を示したりすることで、そのようなプロセスを支援することができるかもしれない。図2は、日本語概念辞書を利用して鮎の方法の「友釣り」の類似語、「あじさい」の方法の「さし木」の類似語を列挙した例である。ここから、鮎をルアーで釣るといったアイデア（このアイデアは実はすでに実現されているのであるが、鮎は一定のサイズになると草食になるため、比較的大きな水生生物を模したルアー、つまり疑似餌に食い付くことは通常ありえない。ここに発想のジャンプがある。鮎のルアー釣りは実はいわゆる友釣りの原理に基づいており、鮎のため



のルアーとは実は喧嘩相手の鮎を模した物である。生きた鮎に対してルアーで喧嘩をしかけると針にかかるという仕組みである。執筆者の内の一人は少年時代、鮎釣りを多いに楽しんだが、このアイデアは晴天の霹靂であった。)、あるいは、あじさいを水耕栽培するといったアイデアを思いつくかもしれない。

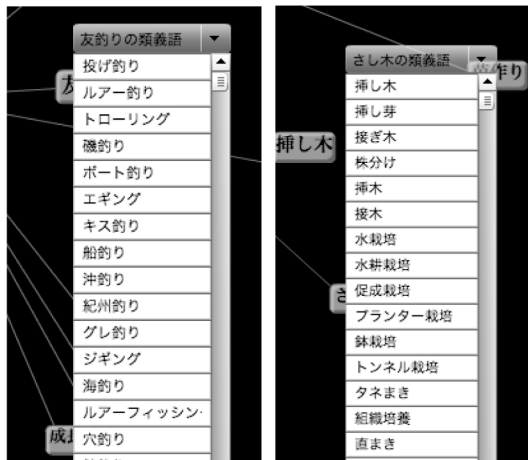


図2 「友釣り」と「さし木」の類似語

また、上位下位関係などは包含する概念をあらわすことから、直接的には新しい手法の発想の支援には役立たないと思われるかもしれない。しかし、属性まで考慮した上位下位関係があるとすると、例えば、「あじさい」 - 「毒のある植物」と「テングタケ」 - 「毒のあるキノコ」という二つの属性を考慮した上位下位関係と、「毒」という属性の共通性から、「テングタケ」を無毒化する方法から「あじさい」を使ったお茶の作成方法を発想することが可能になるかもしれない（日本語概念辞書の上位下位関係はWikipediaから獲得しているため、属性に言及した詳細な上位語が得られることも多い）。この種の推論はいわゆるアナロジーによる推論と呼ばれ、通常の推論にはないアイデアの飛躍があり、既存の特許の範囲

を超えた新しいアイデアということができる。

[アイデア一般化の支援] いったん解決策が得られても、開発者自身がそのアイデアの真の価値を理解しているとは限らず、適用範囲を目下の対象に限定してしまうことなどが起きる。このようなときには、上位下位関係の情報を使ってアイデアの一般化を支援することが有効だと考えられる。例えば、「リンゴ」に対する方法を編み出したとして、「リンゴ」の上位語である「バラ科の植物」にアイデアを一般化してみても、例えば、他の下位語である「梨」でも試してみることで、その手法が「バラ科の植物」一般に有効な方法であることが分かるかもしれない。このように、アイデアの真の価値を見極めて特許出願することで、適用範囲の広いより価値のある特許を得ることができると考えられる。

## 4 今後の課題

現在は、一般のWeb文書から日本語概念辞書を獲得しているが、特許情報処理で有効な支援を行うには、獲得対象の文書を工学・技術分野などに限定して、特許が対象とする分野の概念辞書をより確実に獲得する必要があるかもしれない。しかし、ニーズ発見支援の場面などでは、一般ユーザの「素人考え」のほうが適切かもしれない。そのバランスをいかにとるかが鍵になると考えられる。



## 5 | まとめ

本稿では、現在我々が開発している日本語概念辞書を  
紹介し、それが、特許情報処理にどのように応用できる  
かを考察した。本稿で述べたような、特許検索支援や発  
想支援によって、言語資源・言語処理の社会へのインパ  
クトのある貢献を実現できればと考えている。

