

# 特許情報から用語辞典的なコンテンツを自動抽出し、専門用語辞典を自動作成 最新の技術情報を元にした発想支援や先行技術調査に利用できる有カツールを開発 ウィキペディアやGoogleで見つからない高度な専門用語にも対応可能

特許権は、高度な発明の保護を目的としていますが、日本では年間約40万件の特許が出願され、多様な専門分野に関する発明が蓄積されています。特許に含まれている英知を体系化して活用できれば、学術でも産業でも高い利用価値が期待できます。本研究では、特許情報に発明に関する新語や専門用語が多く含まれていることに着目して、特許情報から用語辞典的なコンテンツを自動構築する技術を開発しました。各見出し語について、説明、分野、英訳、関連語を記載し、関連語どうしをシソーラス（類語辞典）のような階層構造で表現しました。規模と即時性で既存の辞典を凌駕するコンテンツの構築を目指し、構築した辞典コンテンツとシソーラスを特許検索に応用できるようにしました（構築した辞典は現在公開中）。

- 見出し語190万語を収録する用語辞典コンテンツを構築しました。
- 辞典コンテンツの見出し語どうしに関係を付与し、階層的なシソーラスを自動構築できます。
- 公開されている特許情報から辞典コンテンツを構築して、Web情報と特許情報の差異について分析を行いながら、手法やプログラムを拡張しました。
- 検索質問（クエリ）を入力して特許情報テキストを出力するための「検索エンジン」と、検索を円滑に行うための「検索エンジンの操作手段（インタフェース）」で構成されています。
- 特許情報の発明に関する新語や新しい専門用語の抽出結果を基に、ホットになりつつある技術分野が分かり、先端技術調査に役立てられます。

競合技術への強み					
	操作性	自然言語処理能力	類似辞典能力	発想支援	可視化機能
公開特許情報閲覧システム	✗ 初心者には不向き	△ キーワード抽出	✗ なし	✗ なし	✗ なし
本検索システム Cyclone	○ 簡便	○ キーワードだけでなく言葉の説明や関連語も抽出	○ あり	○ あり	○ あり

▲既存のシステムと本技術との比較表

ある技術分野を理解するためには、その分野に属する関連語の集合を知ることが効果的です。本システムを使って、関連語の説明を次々に調べていけば、それらの言葉が発明のポイントであるような特許に自然と到達する確率が高くなります。

## ここがポイント

本研究で重要な技術は、特許情報から専門用語の説明を自動的に抽出する点です。ここでは様々な問題が生じます。まず、ある言葉について説明するときに使われる表現や説明の上手下手は、書き手によって異なります。また、同じ表記で複数の意味をもつ同音異義語があります。例えば、「ハブ」には、ネットワークの集線装置、ヘビ、中心など複数の意味があります。そのため、自分が意図した意味の説明が見つかりにくくなります。

これらの問題を解消するために、本研究は、言葉の説明に使われやすい表現のパターンや統計情報を用いて、説明の候補に点数を付けます。そして、点数が高い説明から順番にユーザに提示します。また、ある言葉の説明によく使われる別の言葉を関連語としてリストアップすることで、情報の絞り込みを容易にします。さらに、言葉の関係をグラフで可視化することによって、ある言葉と関連が強い別の言葉を目で見ながら探すことができます。その結果、ユーザは、自分が意図した意味の説明や関連語を効率的に調べやすくなります。

本研究の用語辞典コンテンツとそれを応用した特許検索インターフェースの実用性を評価するために、特許検索にかかわっているサーチャー、研究者、エンジニア、理工系の大学生・大学院生を中心に合計1000名の被験者を募集し、大規模なアンケート調査を実施しました。被験者の約半数が業務で特許調査を行うことがあると回答を得た方たちであり、被験

業務・研究に役立つか？		
非常に役立つ	145	62.8%が役立つと回答
役立つ	483	
どちらとも	204	
役に立たない	121	
全く役に立たない	47	

▲1000名の被験者へのアンケート調査結果

者に本研究で開発したシステムの一定期間試用を依頼し、複数の設問に回答、評価を提出してもらいました。設問は、普段使用しているツールと比較して本システムで用語の説明が分かりやすかったかどうか、システムの各機能に関する使用感、業務に役立つか等です。その結果、本システムが業務に役立つかという設問に対しては、「非常に役立つ」または「役立つ」と回答した被験者の割合は63%でした。

なお、本システムが主な対象とするサーチャー、研究者、エンジニアほど、本システムが業務に役立つと回答する割合が高くなる、という結果が得られました。本システムは現在、Web上で一般公開（<http://cyclone.slis.tsukuba.ac.jp/>）して、日常的に種々の調査に利用されています。

## ブレイクスルーへの道のり

**2001年**：World Wide Webの情報から百科事典的なコンテンツを構築する本研究プロジェクトの構想段階の提案がIPA未踏ソフトウェア創造事業に採択された。別途、研究していた成果が世界発の多言語特許検索サービス「Patolis-e」を商品化するための研究開発に貢献した。これら2つの研究プロジェクトが後にNEDOの事業として融合されることになる。

**2002年**：2001年の研究プロジェクトがIPA未踏ソフトウェア創造事業に継続して採択され、百科事典的なコンテンツの構築を継続するとともに、当該コンテンツを多面的に活用するための検索インターフェースを開発した。

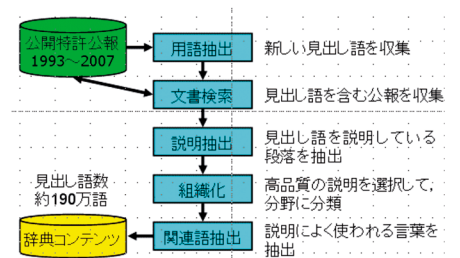
**2003年**：IPA未踏ソフトウェア創造事業の成果に基づいて「事典検索システムCyclone」をWeb上で一般公開した。IPAより「天才プログラマー/スーパークリエイター」を受賞した。

**2004年**：科研費「特定領域研究」の「ITの深化の基盤を拓く情報学研究」において、Cycloneの機能を拡張した。

**2005年**：特許情報の検索に関する研究を進めていたことから、Webだけでなく特許情報からも辞典的なコンテンツを構築できるかもしれないという着想に至り、NEDO産業技術研究助成に応募したところ採択される。日本の特許庁から発行された公開特許公報データの扱いには慣れていたので、研究を始めるための敷居はそれほど高くなかった。

**2006年**：国立情報学研究所主催の評価ワークショップ「NTCIR」で特許検索タスクの運営を行い、米国特許庁から発行された特許データを初めて研究素材として扱った。日本の特許公報に比べると引用文献の提示が厳密に義務付けられているため、引用関係をWebページのリンクと同じように扱って検索できないかと考えた。実践したところ、既存の検索手法よりも高精度で検索できることが分かった。

**2007年**：企業との共同研究を進める中で特許情報処理が主要なテーマになりつつあったことから、本事業の成果を活用できないか模索を始めた。



▲辞典コンテンツ構築システムの概要

## ■サクセス・キー

予想だけで簡単に物事を決めつけず、実際にやってみるまで結果は分からないと自分に言い聞かせて研究を進めました。特許情報から学術研究の対象になることは稀であり、特許情報を研究して何ができるのか？という迷いが生じることもありましたが実際にやってみると予測できない発見がありました。

小規模な実験結果に満足せずに、大規模に研究を展開し、一般に公開するつもりで研究を進めることにより世間の評価にさらされると、研究のレベルでは無視されがちな細かい点にまで注意を払い、新たな発見やノウハウの蓄積につながるがあると考えます。

## ■ネクスト・ストーリー

情報検索、機械翻訳、音声認識のように人間の日常言語を計算機が処理する研究分野では、辞書などの言語資源が「欠かせません。本研究で構築した用語辞典コンテンツをこれらの研究に応用できる水準まで高度化することを目指していきます。

特許情報を用いたCycloneは知財関係者への利用が広がっています。利用者からのフィードバックを受けながら、さらなる実用性の向上を目指していきます。

今回行った研究では、複数の特許情報から抽出した説明を組織化し用語辞典コンテンツ化することによって、個別の特許情報からは得ることができない新たな価値を創造することを試みました。これからも世の中の知的活動に役立つコンテンツの構築を目指していきます。



代表研究者・所属機関・所属部署名・役職名  
藤井 敦 筑波大学  
大学院図書館情報メディア研究科 准教授