

# 特許情報を専門用語辞典として活用するシステム

筑波大学大学院図書館情報メディア研究科准教授  
藤井 敦

## PROFILE

1998年東京工業大学大学院博士課程修了。博士(工学)。現在、筑波大学大学院准教授。2003年IPAから「天才プログラマー／スーパークリエイター」を受賞。自然言語処理、情報検索、音声言語処理の研究に従事。

✉ | [fujii@slis.tsukuba.ac.jp](mailto:fujii@slis.tsukuba.ac.jp)

## 1 はじめに

知的財産権の一つである特許権は、高度な発明の保護を目的としている。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る。特許に内在する人間の英知を体系化し、活用することができれば、今日の高度情報化社会において産業上の価値が高い。

特許には発明に関する新語や専門用語が多く含まれている。本研究は、「言葉」に関する知識に焦点を当て、特許情報を用語辞典として活用することを目的とする。

筆者は、World Wide Webから言葉や事柄に関する説明情報を抽出し、さらに複数の説明情報を組織化することで、百科事典的なコンテンツを自動構築する研究を行ってきた<sup>[1,2]</sup>。さらに、構築したコンテンツに対して、見出し語、関連語、同義語、質問文、関連語マップといった多様な手段によって検索する機能を開発した<sup>[3]</sup>。当該研究成果は、事典検索システム「Cyclone」で一般に公開している<sup>[4]</sup>。

World Wide Webの発展に伴い、Web上のポータルサイト、検索エンジン、ソーシャルネットワーキングなどのツールを使って、様々な調べ物をするのが日常的になっている。しかし、どのような専門用語でもWeb上で調べることができるわけではない。情報技術と関連が低い分野では、Webによる情報発信が盛んでない場合がある。その結果、調べたい専門用語に関する情報がWeb上にないことがある。仮に、専門用語に関する情

報がWeb上にあったとしても、論文や特許に記述されているような詳細な情報が得られるとは限らない。

以下に挙げる専門用語は、Web上では説明を検索することができなかったのに対して、特許情報からは説明を見つけることができた例である。

- 「感光性平版印刷版」

感光性平版印刷版とは、一般に、適当な表面処理を施したアルミニウム、紙あるいはプラスチックなどの支持体の表面に、感光性化合物を含有する感光層を設けたものである。(特開2002-40631)

- 「トラッキング誤差信号」

トラッキング誤差信号とは、対物レンズが光ディスクの半径方向に移動する場合、記録されたピットの中心で0となり、ピットの中心からずれるに従って、値が大きくなる信号である。(特開2001-52347)

以上の背景から、本研究はCycloneの研究で蓄積された技術やノウハウを応用して、特許情報を用語辞典として活用するシステムを提案する。

なお、Webを対象としたCycloneのコンテンツを「事典」と呼んでいることに対して、本研究では「辞典」という言葉を用いる。辞典が専門用語などを主な対象としているのに対して、事典は事件などの事柄も対象としている。本研究では、特許情報に記述された専門用語を対象としているため、辞典という言葉を用いる。ただし、今後、特許情報からも事柄に関する説明を抽出することができることが分かれば、事典という呼び方がふさわしくなる可能性がある。

## 2 辞典コンテンツの構築手法

本研究では、以下の手順によって日本語の公開特許公報から辞典コンテンツを構築する。

### (1) 新語抽出

特許公報から新しい見出し語を抽出する。

### (2) 文書検索

見出し語を含む特許公報を検索する。

### (3) 説明抽出

検索された特許公報の集合から、見出し語を説明している可能性が高いテキストを段落の単位で抽出する。

### (4) 組織化

抽出された複数のテキストに対して、説明としての尤度に基づいてスコアを計算し順位を付ける。さらに分野に分類することで多義や観点による意味の違いを区別する。

### (5) 関連語抽出

見出し語の説明によく使われる言葉を抽出する。

### (6) 要約

説明テキストから、対象の見出し語を説明する観点ご

とに代表文を選択する。その結果、ユーザは短い記述で種々の観点から見出し語の説明を読むことが可能になる。

以上の手順によって、ある見出し語に関する説明と関連語が抽出される。手法の詳細は、藤井 [5] を参照されたい。

## 3 辞典コンテンツの検索手法

本稿執筆当時、公開特許公報1993~2007年発行分を用いて、約190万語の見出し語を含む辞典コンテンツを構築した。構築した辞典コンテンツは、Web用のCycloneと同じ検索インタフェース<sup>[3]</sup>を使用して、見出し語、関連語、同義語、文字列一致、本文一致、関連語グラフによって検索することが可能である。本稿では、見出し語による検索と関連語マップによる検索について具体例を用いて説明する。

図1に、見出し語「トラッキング誤差信号」に対する検索結果を示す。図1の上部には、見出し語に関する「分野」、「関連語」、「複合語」が提示されている。「複合語」とは、関連語のうち、見出し語を含む言葉である。



図1 見出し語「トラッキング誤差信号」に対する辞典コンテンツの検索結果



画面下部には、見出し語に関する説明テキストが2件提示されており、タイトルは抽出元公報の「発明の名称」である。タイトルをクリックすることで抽出元の公報全文を読むことができる。

図1の画面右上にある「関連語マップ」のリンクをクリックすると、図2に示すような「トラッキング誤差信号」を中心とした関連語のグラフが表示される。図2にある長方形のノードは、辞典コンテンツに含まれる見出し語のうち、「トラッキング誤差信号」と関連が強い見出し語である。具体的には、「トラッキング誤差信号」の関連語を取得し、さらに各関連語の関連語を取得して一定数の関連語が取得されるまで再帰的に繰り返す。2つの見出し語どうしが直接の関連語である場合は両者を線で結び、ノード間の重力と斥力によってノードの配置を決定する。

図2のノードをクリックすると、そのノードを中心としたグラフに更新される。また、ノードの上にマウスケールを置くとメニュー項目が表示され、その項目を選択すると、対象のノードを見出し語として図1のような説明の画面に遷移する。すなわち、図1の説明画面と図2の関連語マップを横断することで、ある見出し語を手

掛かりとして関連する言葉の意味を次々に調べることが可能になる。この機能は、特許調査において、調査対象分野の専門用語を調べながら対象の特許公報を検索する場合に有効である。

## 4 考察

Web情報と特許情報からそれぞれ構築したコンテンツの違いについて考察する。前者として、Cycloneで以前より公開しているコンテンツを利用した。

まず、同じ言葉でもWebコンテンツと特許コンテンツでは異なる意味の説明が検索される場合があった。以下に具体例を示す。

- 「コンプライアンス」  
法律遵守 (Web)、変形のしやすさ (特許)
- 「ドレッシング」  
料理 (Web)、研磨 (特許)
- 「ブレイクスルー」  
前進 (Web)、落ち込む現象 (特許)
- 「地下組織」

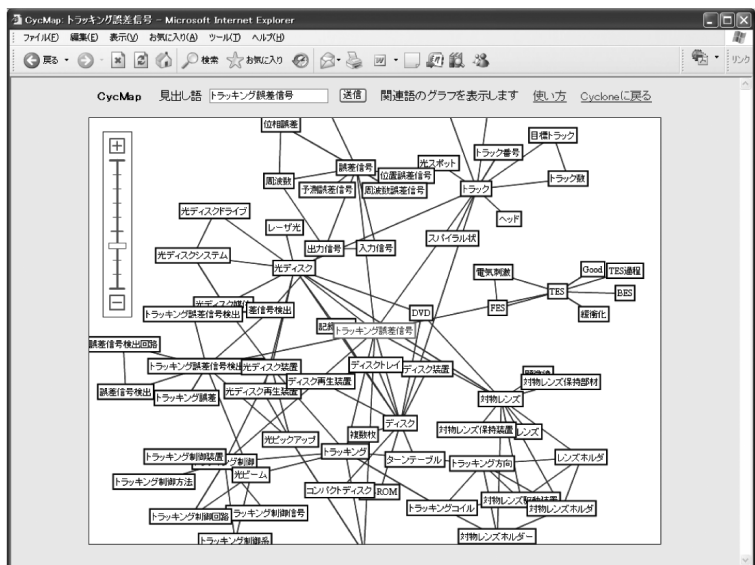


図2 「トラッキング誤差信号」を中心とした関連語マップ

団体 (Web)、根菜の部位 (特許)

さらに、Webコンテンツと特許コンテンツでは説明の詳細度が異なる場合があった。例えば、「オーバーラン」は空気の含有率を表す。特許コンテンツではオーバーランに関する具体的な計算式が検索された。また、「半熟卵」や「温泉卵」に対して、特許コンテンツでは調理時間や加熱温度が数量的に詳しく検索された。

今後、Web情報と特許情報を統合することによって、見出し語の意味や詳細度に関する網羅性が向上し、幅広い情報を提供できる可能性がある。

pp.31-36, 2004.

[4] <http://cyclone.slis.tsukuba.ac.jp/>

[5] 藤井敦. 特許情報を用いた辞典検索システム. 情報処理学会研究報告, 2008-FI-91, pp.9-15, 2008.

## 5 おわりに

特許情報から用語辞典的なコンテンツを自動構築し、多様な手法でコンテンツを検索するためのシステムを提案した。特許情報から構築した用語辞典コンテンツと検索システムもCycloneのWebサイトで公開中である。

### 謝辞

本研究は、NEDO「産業技術研究助成事業」の助成で行った。

### 参考文献

- [1] Atsushi Fujii and Tetsuya Ishikawa. Summarizing Encyclopedic Term Descriptions on the Web. Proceedings of the 20th International Conference on Computational Linguistics, pp.645-651, 2004.
- [2] 藤井敦, 石川徹也. World Wide Webを用いた事典知識情報の抽出と組織化. 電子情報通信学会論文誌 D-II, Vol.J85-D-II, No.2, pp.300-307, 2002.
- [3] 藤井敦, 伊藤克亘, 石川徹也. Webマイニングによる事典的コンテンツの構築と多様なアクセス手法. 電子情報通信学会技術研究報告, DE2004-6,