

# 特許文献を対象とする意味関係の抽出 -思考支援システムを目指して-

石川大介†

石塚英弘‡

宇陀則彦‡

藤原謙\*

図書館情報大学大学院情報メディア研究科†

図書館情報大学図書館情報学部‡

工業所有権総合情報館\*

## 1 はじめに

本研究は、主に専門用語を中心として用語間関係の構造化を行ない、それを基盤の知識として推論などを行なうことによって、人間の研究開発における思考を支援するシステムの構築を目指している。この用語に基づく知識体系を構築するために、さまざまな用語間の意味関係の抽出を行なう。

この意味関係の抽出のために、まず特定の意味関係を表現する特定の表現パターンを、対象となる媒体及び分野から調査した。本研究では、この調査の段階において、既存の意味関係の知識を利用して、意味関係の表現を調査した。

## 2 意味関係の抽出

専門分野の文献には、さまざまな意味関係が記述されているため、ここから意味関係を抽出できる可能性がある。その方法として、例えば特定の意味関係を表現する特定の表現に注目し、それを利用して抽出する方法が挙げられ、我々もそれを行なった [1][2]。[2]の研究では、光学材料ハンドブックを利用して、非線形光学材料情報の意味関係の抽出を行なっている。この中で、括弧を使った同値関係の表現や、「AにはBがある」「AとしてBがある」などの上下関係（階層関係）の表現に注目し、抽出を行なっている。これらの方法では、特定の意味関係を表す表現パターンを探し出す作業が重要と考えられる。

特定の専門分野では、用語の意味関係が既に明確に定義されている。このような既存の意味関係は、特定の意味関係を表す特定の表現を探すために有効と考えられる。そこで、本研究ではビタミンやアミノ酸における既存の意味関係を使って、特許文献を対象として意味関係を表す表現のパターンを調査した。

## 3 特許文献の特徴

特許文献を知識の抽出対象と考えた場合、以下の点が特徴として挙げられる。

- 自然科学を利用した知識
- 産業などに応用された知識

Extraction of semantic relationships in patent documents  
-Toward thinking support system-  
Daisuke ISHIKAWA†, Hidehiro ISHIZUKA‡, Norihiko UDA‡, Yuzuru FUJIWARA\*  
Graduate School of Information and Media Studies, University of Library and Information Science†  
Faculty of Library and Information Science, University of Library and Information Science‡  
National Center for Industrial Property Information\*

- 実現可能な知識

そのため、特許文献から得られた知識は産業などに応用可能な思考支援システムに役立つと考えられる。

また、特許文献を利用する利点は、文献から得られた知識に以下の書誌情報から次の情報が付与される点である。

- 出願日 → 新しさ
- 出願人（発明者） → 誰の考案か
- 国際特許分類 (IPC) → 適用される分野

## 4 実験

研究利用可能かつ機械可読な特許文献に、2002年 NTCIR ワークショップの NTCIR-3[3] にて配布された特許コーパスがある。この中から、1998年度版の Japio 出願抄録 (日本語) の約 34 万件を使用した。これから、特許の分類コード IPC: A61K「医薬用、歯科用又は化粧品用製剤」を用いて 5205 件を取り出し、これを対象テキストとした。以下、この対象テキストを対象に実験を行なった。

### 4.1 同値関係

同値関係では、例えばビタミンを例にすると、ビタミン C とアスコルビン酸、ビタミン E とトコフェロールはそれぞれ同じ意味とされている [4]。この同値の両方の用語が使われているテキストを対象テキストから検索した。その結果、例えば以下の表現があった。

- 好ましくは ビタミン C (アスコルビン酸) である ~

同値関係にある用語を A と A' とした場合、この表現は「A(A')」と表すことができる。この表現パターンを使った他の同値関係の用語を調査した。以下、その結果を示す。どれも、A と A' は同値関係にある。

- 特に メチシリン耐性ブドウ球菌 (MRSA) ~
- ヒト免疫不全ウイルス (HIV) の細胞への感染を ~
- ~または ドコサヘキサエン酸 (DHA) である。

### 4.2 階層関係

典型的なアミノ酸として、アラニン、アルギニン、アスパラギン、...、バリン、チロジンの 24 種類がある [4]。この場合、アミノ酸を親とし、アラニン等を子とする階層関係がある。「アミノ酸」とアラニン等、個々のアミノ酸の両方の用語が用いられているテキストを対象テキストから検索した。以下、その結果を示す。

1. アミノ酸としては L-プロリン、L-イソロイシン、L-ロイシン、L-アラニン、L-スレオニン、L-リジン が好ましく、～
2. グリシン、L-アラニン、L-バリン、D-イソロイシン、D,L-リジン 等の アミノ酸、及び～
3. 必要に応じて アミノ酸 (グリシン 等) とを配合し、～

これらの表現の中で、我々は3.の括弧を使った表現パターンに注目した。この表現パターンを使った他の階層関係の用語を調査した。以下、その結果を示す。

- この薬剤は、内服剤(カプセル、錠剤 等) 外用剤(軟膏 等) の形態で、～
- 例えば 有機溶媒 (メタノール 等) または～
- 油性食品 (バター、マガリン 等) 中に～

### 4.3 物質とその作用の関係

ビタミンにはいくつかの作用がある。このような物質とその作用の関係を調べるため、「ビタミン」と「作用」で検索し、この関係の表現を調査した。その結果、例えば以下の表現があった。

- 免疫調節作用 を有する 活性型ビタミンD誘導体 の 合成中間体 等としてし有用な、～

これは、物質を A、その作用を として場合、「 を有する A」と表せる。この表現を使った他の物質とその作用の関係について調査した。以下にその結果を示す。

- 又、脱臭作用 を有する ヤマブシタケ抽出物 を得るための抽出溶媒としては～
- 線維芽細胞増殖促進作用 を有する A-モンド 等の植物抽出物を～
- ～と特定の UV-Bフィルタ-作用 を有する トリアゾール化合物 とを特定の割合で～

## 5 考察

今回の実験では、同値関係や階層関係を記述する括弧の表現に注目した。しかし、括弧には想定とは異なる使われ方が確認された。以下にその例を示す。

- 化学式の説明：式 ( $R_1$ 、 $R_2$  は独立的に H、 $C_{1 \sim 8}$  アルキル、アリール、アルケニル等； $n$  は 2～4) で表される～
- 補足：セリン (誘導体)、ニコチンアルコール (塩)
- 単位：(モル) (重量)
- 箇条書き：(1)～(2)～(3)～

こうした表現に対応するためには、辞書などを利用し、括弧の中の用語が辞書に登録されている用語であるか確認するなどして対処することが考えられる。

物質とその作用の関係を表す「 を有する A」の表現は、さまざまな関係が記述される。そのため、例えば、には語尾が「～作用、～効果」の用語だけを限定し、Aには辞書に登録された用語から選択されるようにするなどの工夫が考えられる。

## 6 他の分野について

今回の実験で得られた特定の意味関係を表す表現について、他の分野にも同様に使用されているかを調べた。繊維工学の分野 (IPCコード D06M：繊維の物理的、化学的、その他の処理方法) を対象に調査を行なったのでその結果を以下に示す。

- 同値関係：～を布帛3上に 積層 (ラミネート) させて形成しており、～
- 階層関係：～マルチフィラメント糸を 溶解処理 (アルカリ減量 等) することにより得られる。
- 作用の関係：光触媒作用 を有する 酸化チタン を繊維に特定量含有せしめることにより、～

このように、他の分野でも同様の表現が存在し、それで意味関係を表現していることが分かった。

## 7 まとめ

今回の研究では、意味関係を抽出する前に必要となる有効な表現パターンの調査を、既存の知識を利用して行なった。今後は、この結果に基づいて意味関係を抽出し、別の方法で求めた因果関係 [5] との統合を図る。そして、統合された知識に基づく推論により、思考が支援されるシステムの開発を目指す。

## 謝辞

本研究において、国立情報学研究所で作成された NII-NACSIS コレクションの NTCIR-3 を使用しました。深く感謝いたします。

## 参考文献

- [1] Hikomaro Sano, Yuzuru Fujiwara: Syntactic and semantic structure analysis of article titles in analytical chemistry, Journal of Information Science 19, pp.119-124, 1993
- [2] 宇陀則彦、石川雅弘、山本毅雄、藤原謙：非線形光学材料情報における意味関係の抽出と情報の構造化、第3回研究報告会講演論文集、pp.67-76、情報知識学会、1995
- [3] NTCIR : <http://research.nii.ac.jp/ntcir/index-ja.html>
- [4] Earl Mindell 著、丸元淑生 訳：ビタミン・バイブル、小学館、1993
- [5] 石川大介、石塚英弘、宇陀則彦、藤原謙：特許文献を用いた因果関係に基づく知識構造化の試み、情報処理学会研究報告 2003-FI-72、pp.61-67、2003