

# 専門用語の専門性判定に関する一考察

大学共同利用機関法人情報・システム研究機構 国立情報学研究所学術コンテンツサービス研究開発センター特任研究員 内山 清子

## PROFILE

1997年慶應義塾大学大学院政策・メディア研究科修士課程終了、1998年学術情報センター COE 研究員、2003年スタンフォード大学 CSLI 訪問研究員、2005年学術博士（慶應義塾大学）、2005年～2008年慶應義塾大学特別研究教員准教授、2009年6月より現職。組織 大学共同利用機関法人 情報・システム研究機構。2009年度から特許産業日本語委員会委員。

✉ kiyoko@nii.ac.jp



## 1 はじめに

特許文書、学術論文、技術報告には専門用語が数多く含まれている。しかし、実際には専門用語、一般語の区別は明確ではなく、各分野の専門家であれば明らかに専門用語であると判断できるが、新しい技術に関する用語が出てきた場合、共通して専門性を認識するまでには時間を要する。また、ある分野の初心者にとってその単語が専門用語であるのか、そうではないのかの判断が難しい。専門用語だとわかったとしても、どの程度の専門性であるのか、たとえばある特定分野における基礎的・必須用語か、基礎的用语の知識がなければ理解できない専門性の高い用語であるのかなどの段階はわからない。これらの段階がわかれば、難解な用語であっても、基礎的・必須用語を用いた説明を加えることにより理解を容易にすることができる。分野個別の専門用語は専門用語辞書で調べるのが先決であるが、技術進歩が速い分野では辞書のメンテナンスが追いついていない。特に特許文書の場合は、最先端の技術を申請するものであるため、一般的な学術論文で使われている専門用語だけでは表現できない用語が出てくる。このような背景の中で、技術文書に含まれる用語の特徴とそれらを扱うために、今後必要となる指標について記述する。

## 2 専門用語の出現傾向

まず、本文における専門用語の定義は、「特定分野において特徴的な独自の事物、事象について他の分野あるいは一般語とは異なる名称」とする。中には同じ名称であっても全く違う事物、事象を指すことがあるが、それも専門用語とする。専門用語は特定分野の専門家が作成した用語集に掲載されれば公に専門用語と認定されることになる。しかし、用語集は一度作ってしまうと新しい用語を追加して再度用語集として出版することはあまりない。現在では、Web上で用語集を作成して、日々用語を更新できる仕組みが出来つつある。Web上の用語集は新聞や雑誌に多く取り上げられる言葉を収集して、重要と思われる語を用語集に掲載している。そのため特許や最先端の技術報告、学術論文などに含まれる用語や、マイナーな分野の用語は収集されにくい。つまり、用語集に掲載されていない用語は公に専門用語であると判断することが難しい状況である。用語集にはまだ掲載されていないが、その分野では当たり前に使っている用語、最先端であるが今後重要になってくる用語にどのように対応していくかが、問題となっている。

そこで、学術論文を例にあげ、特許文書への応用や今後の展開について述べる。まず学術論文の場合、著者が「著者キーワード」を著書論文に付与することによって

いる。この著者キーワードは特許文書におけるFタームに相当するが、Fタームのようにあらかじめ規定されていないため、記述における自由度が高い。著者キーワードの出現傾向を情報処理学会刊行誌論文データから41の研究会、1993年から2005年までの13年間の著者キーワードの出現頻度を調べた結果、多くの著者が共通して使用するキーワードと著者だけが独自に使うキーワード（頻度1、全体の65%）に2極化していることがわかった[1]。また、著者キーワードは平均3から5個付与するが、広い分野における共通のキーワード、比較的狭い特徴的な用語、著者独自のキーワードのような付与をしていた。

次に著者キーワードの出現分布を出現分野と出現年数で分析することで専門用語の特徴を調べた。まず、出現分野については、情報処理学会は(1)コンピュータサイエンス、(2)情報環境領域(3)フロンティア領域の3つに分かれているため、この領域を分野とする。その分野に出現した年度と頻度を数えて、出現傾向を分析した。分析の結果、(1)「インターネット」「アルゴリズム」等の10年以上継続的に全分野に共通して数多く出現(2)「ウェーブレット変換」等の継続的に全分野に数少なく出現(3)「形態素解析」「構文解析」などの特定領域に10年以上数多く出現(4)「句構造規則」「nグラムモデル」等の特定領域に断続的に数少なく出現するという大きく4つの傾向が見られた。専門性の観点からは(4)が専門性が高く、(1)は専門性が低く、より一般的な用語であると考えられる。出現分野、出現年度の傾向によりある程度専門性を推測できるのではないかと考えられる。

### 3 特許文書への適用

学術論文において2で示したような傾向があることから特許文書でも用語の出現傾向を分野や年度の視点か

ら調べることにより、専門用語の判定がある程度できるのではないかと考える。特許文書の場合、共通して使用する用語が多様なため、用語表記の統一や、同じ動詞に共起する名詞の出現傾向をみるなどの工夫が必要になってくる。特許文書に付与されているFタームを領域タグとして、Fタームと特許文書中の用語との対応づけを行うことで広い分野における専門用語であるのか、より狭いサブカテゴリの専門用語であるのかを判定することが可能である。

さらに、専門用語が文中で用いられる語のパターンや共起語に着目（「AなどのB」という表現ではAはBの一種であり、Bは基本的な用語と考えられるような文脈情報や専門用語を構成している形態素の数（形態素の数が長くなると専門度が増す）や、その特徴に関する語彙情報を用いることにより専門性の指標になると考える。

## 4 今後の課題

専門用語の専門性を判定するための情報についての分析と考察に基づいて、専門性の指標を示し、その指標を用いて平易な用語で記述することにより、読みやすい特許文書の記述に貢献することが可能だと考える。

謝辞

情報処理学会刊行誌掲載論文本文データに関して、研究利用することを許諾していただいた、社団法人情報処理学会に感謝いたします。

[1] 内山清子、鈴木崇史、相澤彰子、「専門用語の専門度の指標に関する一考察」、言語処理学会第16回年次大会(NLP2010),pp.571-574, 2010.