

検索精度向上への取り組み

財団内研究所の活動について

財団法人工業所有権協力センター 研究所
総括研究員

居島 一仁

PROFILE 平成19年より現職



1 はじめに

財団法人工業所有権協力センター（IPCC：Industrial Property Cooperation Center、以下「財団」と表す。）は、「工業所有権に関する手続き等の特例に関する法律」（平成2年法律第30号）に基づく登録調査機関として、特許出願等への分類付与、特許出願に関する先行技術調査（検索）などの業務を行っている。

飛躍的に増加する検索対象文献数及び年々増え続ける技術高度化負担への対応や今後とも増大が予想される検索業務量等への対応は、現在財団が解決すべき喫緊の課題となっている。

こうした課題への解決策として、財団では、主席部員の採用促進、各種システム開発による業務負担の軽減を図るとともに、研究所において、業務効率化に資するテーマを設定し、研究を行っている。

2 研究活動について

財団に付置された研究所では、工業所有権情報の分類及び分類を用いた検索システムについての調査、研究及び開発に関すること、工業所有権情報の分類及び分類を用いた検索システムについての調査、研究及び開発に関すること等を行ってきたが、今までの研究結果に基づき、検索業務（特にテキスト検索）を行うには、シソーラス辞書の充実化が重要であるとの認識から、近年は

シソーラス辞書の充実化に研究の重点を置いている。

3 IPCCシソーラスについて

特許文献の検索として、FIやFタームなどのコード記号による検索手法と言葉を利用したテキスト検索があるが、テキスト検索は、技術進歩が急速な分野に有効な検索手法として多くの部門で利用が拡大している。

一方、言葉による検索は同一の対象を表現する技術用語や言い方が複数あるため、類義語の問題を常に抱えている。同一内容を異なる文献では異なる表現で記述している場合がよくあり、出願人が違つてこのことはより顕著となる。

そのため、検索者は複数の類義語を論理和で結合した論理式を作成して検索することで、検索結果に漏れが生じないように努力することが必要になる。

しかしながら、重要な類義語に気づかずに類義語を指定し忘れると、検索結果に漏れが生じてしまう。特に経験の浅い検索者、あるいは新規の分野を担当する検索者は、このような事態を招く危険が高いといえる。

類義語の指定漏れを防ぐ方法として類義語辞書の利用があるが、財団では、財団職員が検索業務を行う中で蓄積してきた情報を集約し構築した独自の辞書としてIPCCシソーラス辞書を開発し、その構築方法等について研究してきた。

IPCCシソーラス辞書は、広義のシソーラスとして（イ）異表記（「コンピュータ」と「コンピューター」、

「ガラス」と「硝子」、「組合せ」と「組み合わせ」など、単語の表記上のバリエーションが異表記。）、(ロ) 類義語（「コンピュータ」と「電子計算機」、「携帯機器」と「モバイル情報端末」、「レポート」と「報告書」など、単語が表す概念は完全に一致するわけではないが、特定の分野あるいは特定の文脈のもとで意義の類似する言葉と見なすことができる語。）、(ハ) 上位語・下位語（「弾性体」と「ゴム」、「回路」と「電子回路」、「材料」と「耐火材料」など、概念的に上位あるいは下位の関係にある語。）、(ニ) 関連語（準類義語：上記（イ）から（ハ）のいずれにも該当しないが、「梅雨」と「湿気」、「切换え」と「選択」、「押下」と「指示」、「落雷」と「停電」など、単語が表す概念あるいはその概念から演繹または敷衍した概念になんらかの共通が認められる語。）が含まれたものとして構成されている。

4 シソーラスの有効性

財団が特許庁から受注し、平成18年10月～11月半ばに納品した検索報告書を調べたところ、全検索論理式中の約6割がテキスト検索式を用いており、テキスト検索式において類義語展開が行われているものが約7割（全検索論理式中では約4割）と高く、類義語展開は必須かつ一般的であった。

このようなテキスト検索を行う際には、検索式において、十分な類義語展開が行われることが検索品質の確保に必要であるが、前記シソーラスデータを利用することにより、類義語展開の支援・確認等を簡易に行えること、主席部員間（特に新人主席部員と検索経験豊富な主席部員）でシソーラスデータを共有できる等の利点を有していることから、前記シソーラスデータは財団内で有効であると考えられる。

財団内でテキスト検索を行う際に、IPCCシソーラス辞書を参照することで、類義語の指定漏れを最小限に抑えることが可能となり、これにより検索の質を向上する

ことができる。

現在、IPCCシソーラス辞書は、(イ) 2,900を超えるテーマコード別に各テーマ特有のシソーラスにより構成されたテーマ別辞書（総登録語数：約615万語）と(ロ) テーマ別辞書内の類義語をある特定のルールの下統合した統合辞書（総登録語数：約272万語）により構成されており、テーマ別辞書は各テーマ内を検索する際に有効であり、一方、統合辞書はテーマに左右されることなく、全テーマにまたがった類義語が登録されているため、テーマをまたがった検索を行う場合やより多くの類義語を参照したい場合に有効である。

5 今後の研究について

財団では、業務効率化の目的で、上記シソーラス等に関して、抽出されたシソーラス情報を利便性の高いデータに加工する等の研究を継続している。

なお、重要と考える研究成果について、特許出願を行っている。

一方、特許庁の推進する最適化計画において実現される次期検索系システムの開発動向に留意しつつ、検索業務効率化に資する調査・研究を行うとともに、公益法人として広く公益に資する研究成果の蓄積や公表のあり方等について検討することも財団の今後の課題といえる。