

特許の先行技術調査における集合知の循環

橋田 浩一 (産業技術総合研究所)・藤井 敦 (東京工業大学)・岩山 真 (日立製作所中央研究所・東京工業大学)・

橋本 泰一 (東京工業大学)・乾 孝司 (筑波大学)・難波 英嗣 (広島市立大学)

1. はじめに

本論文では、調査者の集合知に基づいて意味情報を集約し共有することにより、特許の先行技術調査等における業務全体の効率と品質を持続的に向上させる方法について論ずる。

各調査者は、審査等の対象である特許(本願)における各請求項とその実施例等との対応関係に関する注釈を作成し、本願の読解(下図の①)や他の調査者等とのコミュニケーションに用いる。先願に同様の注釈が付与されていればそれは検索のヒントとして使うことができ(下図の②)、また先願を迅速に読解する助けにもなる(下図の③)。この注釈と検索の履歴から辞書を自動生成し(下図の④)、この辞書と注釈と検索履歴とを検索質問の生成に使いながら(下図の⑤⑥⑦)修正・拡張する。このように注釈と検索質問と辞書を通じて調査者の集合知を循環させることにより、特許の先行技術調査の生産性を持続的に高めることができると考えられる。

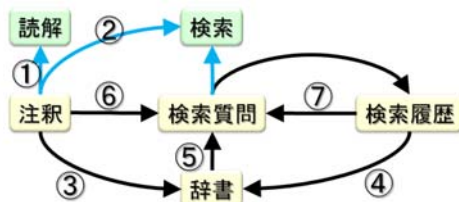


図1: 集合知を用いた特許の先行技術調査

以下では、このアーキテクチャに含まれる要素技術とその統合について述べる。

2. 注釈

特許文書の中核をなすのは請求項であるが、請求項は請求の範囲を広くするために抽象的に書かれていることが多く、必ずしもわかりやすくはない。一方、請求項の実施例は具体的に書かれることが多いので、対応する実施例を参照することによって請求項の内容が理解できると考えられる。したがって、請求項と実施例との対応関係に関する注釈を特許文書に付与することによって、特許文書の理解が容易になるだろう。

このような注釈の付与は、特許の先行技術調査の調査者が本願を理解しようとする際に何らかの形でしばしば実際に行なっている作業である。ゆえに、その形式を

標準化し共有・再利用可能にするとともに、従来の作業に比べて余分なコストを伴わないような注釈付与のインタフェースを開発することによって、全体としての生産性の向上が期待できる。

調査者からのヒアリングを含む著者らの検討によれば、この対応関係の項(argument)としては、請求項においては構成要素、実施例においては段落が、大きすぎず小さすぎないという意味で適当と考えられる。ひとつの請求項全体を対応関係の項とすると、概念の広い請求項の場合には対応先が長くなって注釈の意義がほとんどなくなってしまふ。一方、対応関係の項を構成要素よりも小さな単位とすると、注釈のコストが過大になる可能性が高い。

また、対応関係の種類としては、対応先の段落は対応元の構成要素の「詳細化」である場合が多いが、「効果」であることもある。他の関係もあるか否かは調査中である。「詳細化」は後述のように検索質問の生成に有用であり、また「効果」は「詳細化」とは別の意味で請求項の理解に役立つので、対応関係の種類を区別して注釈を付与するのが望ましいと思われる。

以上のような考察に基づいて、各請求項とその実施例等との対応関係に関する注釈を表現するフォーマットを定式化し、それに基づいて注釈を手で付与するためのインタフェースを開発することを考えている。

3. 辞書

検索履歴から辞書を自動生成する技術については[2][5]を参照されたい。請求項(の中の構成素)と実施例の対応関係がわかれば、請求項に含まれる語 W とその請求項の実施例に含まれる W の下位語との関係が高い精度で抽出できると考えられるので、上記のような注釈も用いることによって辞書の精度を向上させることができるはずである。これらの方法を統合することにより、精度の高い辞書を生成する技術を開発できるだろう。

さらに、オントロジーを共同編集する基盤にこの辞書を格納し、別種の集合知に基づいて、オントロジーとしての辞書を拡張・改良することが考えられる。Fタームは一種のオントロジーのようなものであるから、このオントロジーの一部として保守・拡張することができるだろう。Fタームは公開コンテンツだから、この集合知には、特許の先行技術調査の調査者に限らず、研究者や技

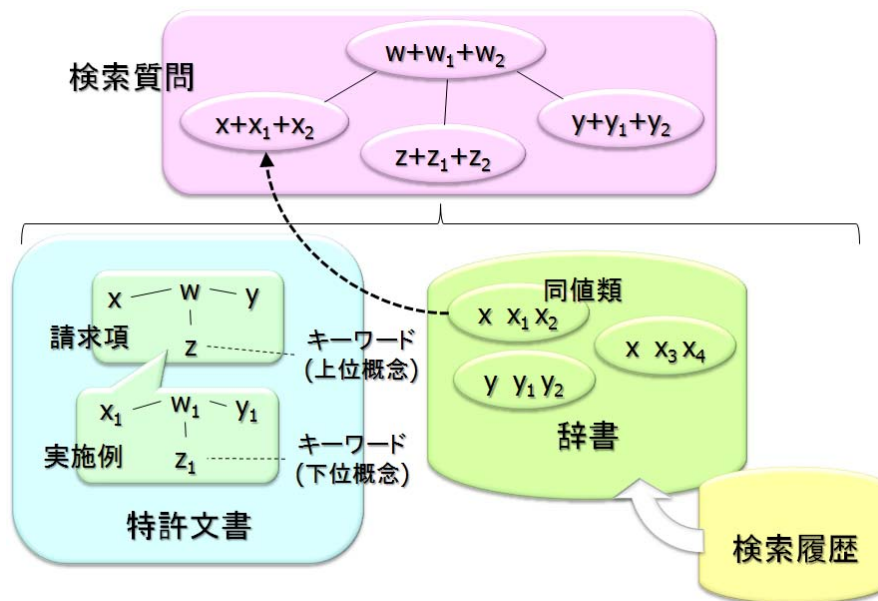


図 2: 注釈と辞書からの検索質問の生成

術者を含む不特定多数の利用者が参画できるだろう。

4. 検索

辞書と注釈と検索履歴から検索質問を自動生成し、これを必要に応じて人手で修正しながらインタラクティブに検索を行なうインターフェースの開発を予定している。その際の各検索は、先願に付与された注釈も参考しつつ何らかの近似的な意味構造の照合を伴うものになるだろう。

検索質問の生成の方法を図 2 に示す。すなわち、特許文書において請求項とその実施例が注釈によって対応付けられており、請求項の中の用語(たとえば x)とそれに対応する実施例の中の用語(請求項の中の用語またはその下位語; たとえば x_1)との関係が検出できれば、それを辞書(シソーラス)の中の用語の同値類(たとえば $\{x, x_1, x_2\}$)で拡張して請求項の意味構造に代入することによって検索質問を自動生成する。請求項が構成要素に区切られていれば、意味構造は高い精度で求めることができるだろう。

5. おわりに

特許の先行技術調査において集合知を利用する方法について述べた。注釈等から検索質問を生成する技術 [2][4] や意味構造を用いてインタラクティブに検索質問を改訂する技術 [1] は既知であるが、この方法は、それらを集合知の循環に組み込んで先行技術調査全体の生産性を持続的に向上させることを狙う。

以上の技術を統合した図 1 のようなシステムを特許の先行技術調査の現場において利用できる環境を整え、日

常的にシステムを利用し評価しつつ注釈と検索履歴を蓄積できるようにすることにより、このシステム全体(人間系を含む)を持続的に改良できるような体制を構築したい。

謝辞

本研究は IPCC (工業所有権協力センター) の特定目的寄付金「特許の先行技術文献調査に関する次世代技術とその利用技術に関する研究」によって行なわれている。

参考文献

- [1] 橋田 浩一・宮田 高志 (2006) 類義性計算方法、類義性計算プログラム、類義性計算プログラムを記録したコンピュータ読み取り可能な記録媒体。特許第 3856388, 産業技術総合研究所・科学技術振興事業団。
- [2] 乾 孝司・難波 英嗣・橋本 泰一・藤井 敦・岩山 真・橋田 浩一 (2011) 最大クリーク探索に基づく特許検索履歴の統合。言語処理学会第 17 回年次大会。
- [3] 出羽 達也 (2000) 文書検索方法および文書検索装置。特開 2000-339342, 株式会社東芝。
- [4] 小西 一也・北内 啓・高木 徹 (2005) 類似特許明細書検索システム及びその方法並びにプログラム。特開 2005-234868, 株式会社エヌ・ティ・ティ・データ。
- [5] 難波 英嗣・竹澤 寿幸・乾 孝司・岩山 真・橋田 浩一・橋本 泰一・藤井 敦 (2011) 特許検索履歴を用いたシソーラスの自動構築。言語処理学会第 17 回年次大会。