

大規模シソーラス辞書を適用した 特許対比支援システム

NTTデータ 技術開発本部
ビジネスインテリジェンス推進センター

近藤 功

PROFILE

2002年3月北海道大学大学院修士課程修了。同年(株)NTTデータ入社。2004年NTTサイバーソリューション研究所勤務を経て、2007年より現職。テキスト処理技術の研究開発に従事。



NTTデータ 技術開発本部
ビジネスインテリジェンス推進センター

小西 一也

PROFILE

1997年3月茨城大学大学院修士課程修了。同年より、NTTデータ通信株式会社(現株式会社NTTデータ)入社、2000年NTTサイバースペース研究所勤務を経て、2003年より現職。特許文書解析技術の開発に従事。



東京大学史料編纂所
前近代日本史情報国際センター・特任教授

石川 徹也

PROFILE

1971年3月慶応義塾大学大学院修士課程修了。富士フイルム(株)足柄研究所、図書館短期大学、図書館情報大学、筑波大学などを経て現職。現在、歴史知識学の創成研究に従事。筑波大学名誉教授。工学博士。2008年度から特許産業日本語委員会委員。

✉ ishikawa@hi.u-tokyo.ac.jp



1 はじめに

審査における特許調査の効率化に向けた支援の一つとして、本願と引例文献候補の対比システムを紹介する。

対比することは、特に審査における特許調査で、非常に重要な業務である。この対比業務には、類似判断を適切に行うために、比較元(本願側)の発明特定事項単位に、比較先(引例側)に対して、その記載の有無を確認する必要がある、という特徴がある。そのため、明細書の部分対比を容易にする対比支援が重要である。

一般的な文書の対比は、元々同一の文書の改変部分を見つけ出すことを基本としている。ほぼ同内容の文書の違い(差分)の発見をサポートすることが、対比支援ツールに求められる機能となっており、文書を指定するとその差分を履歴として保持および表示できるツール(システム)が実際に市販されている。

しかし、特許調査における対比は、基本的に元々別の著者によって書かれた別の内容の文書同士を比較する作業であり、その目的は差分を見つけ出すことではなく、逆に類似部分を見つけ出すことを基本とする。したがっ

て特許調査の対比支援ツールに求められる機能は、本願と引例候補の記述の中のお互いの類似部分の発見にある。

提案する対比システムは、検索に用いることが多いシソーラス辞書を、文書対比を目的に適用することで、従来よりも精度の高い対比を可能としている。

2 特許調査の対比支援方法

2.1 特許調査における対比の問題

特許調査における対比の問題は、一方の文書データに記述された内容が他方の文書のどの部分に対応するのかを把握することが困難な点にある。

記載内容の理解を支援する一般的な方法として、ユーザが指定するキーワードを強調表示する方法がある。また、従来技術を拡張したより高度な支援方法も多数検討されている。その代表的な方法に、ユーザが指定した複数のキーワードと色によりテキストを強調表示すると共に、指定キーワードがテキスト全体のどの位置に存在するのかを俯瞰するスペクトルバー表示を備えるインター

フェース技術がある。

しかし、いずれの方法も、以下の2点の問題を解決するには至っていない。

(1) 一般に長大なテキストデータである特許文献同士を対象に類似箇所をの対比を行うには、情報の一覽性に欠ける。

従来方法は、注目する記載内容の発見に役立つ機能であるが、本願の発明特定事項単位に対応する記述箇所の候補を見つけて、総合的に本願の類似先願になり得るのかを判断するという用途を考えると、一覽性に欠ける。

(2) 意味として同様な内容が、別々の語句で記載されている場合に、その記載箇所を的確に把握することは困難である。

この問題に対しては、指定するキーワードの類義語を多く指定することで、運用的な対処が可能である。

2.2 提案する特許対比支援システム

2.1 で述べた問題解決にむけて、(1) の問題については、本願と引例文献候補のスペクトルバー表示を並列配置するインターフェースの導入を、(2) の問題について

は、大規模シソーラス辞書の適用を行った。

(1) 本願・引例の並列スペクトルバー表示

長大な特許文書全体を俯瞰できるスペクトル表示を、本願と引例候補文書の両方に適用し、(図1)に示すように左右に配置するユーザインターフェースを導入する。並列配置することで、本願側で注目する記述箇所の特定と、それに対応する引例側の記述内容の有無を確認する作業が連携できる。そのため、本願の発明特定事項単位に繰り返し行う部分対比作業の負荷を軽減できる。また、本願側にもスペクトルバー表示を備えており、本願側で注目する発明特定事項に対応する発明の詳細な説明部分も容易に確認できることから、類似判断が容易になる。

構築した特許対比支援プロトタイプシステムは、(図1)左側に示すように、本願の注目する記述箇所から記述内容を特徴付けるキーワードを、色と共に指定すると、テキスト本文およびスペクトルバー表示部を指定色で強調表示する。スペクトルバー表示部の1行が、テキスト本文の部分テキスト(数文字)に対応し、スペクトル1行の色分けは、部分テキスト内における指定キーワードの有無を表す。また、スペクトルバー表示部の上部は、テキスト本文の先頭部分に対応する。このスペク

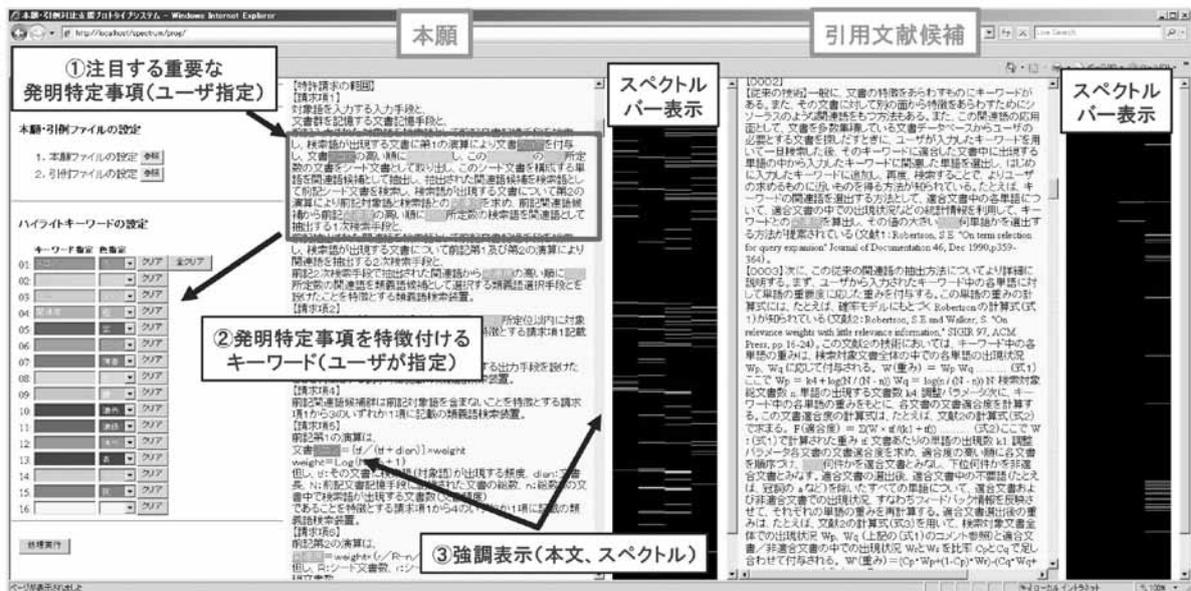


図1 本願・引例候補文献の対比支援プロトタイプシステムの画面



トルバー表示を用いることで、ユーザが指定するキーワードの有無をテキスト全体で俯瞰しながら確認できる。

(2) シソーラス辞書の適用

検索にシソーラス辞書を適用する際の課題の一つに、複数の類義語候補の中から検索に有効な類義語を選択することが困難であるという課題がある。

この課題を踏まえ、特許対比支援システムにシソーラス辞書を適用する場合は、本願から抽出した特徴的なキーワードのうち、引例側に現れないキーワードのみを対象に、引例側に出現するシソーラス内の登録語を拡張する。拡張する候補語を引例文書内の語句に限定することで、ノイズが減り、対比に有効な語句を効率良く提示できると考える。(図2)に、シソーラス辞書に基づくユーザが指定したキーワードの拡張イメージを示す。

本システムでは、特許対比に用いるシソーラス辞書として、特許文献の筆者が実際に言い換えた表現を収集した事例集を辞書として適用した。構築辞書は、525万3617件の公開特許公報(1993-2007年15年分)から独自のルールに基づき抽出した語句515万7471件と、各語句の関連度から成る。

対比に用いた辞書には、膨大な登録語数を有するだけでなく、特許特有な表現を含むという特徴がある。そのため、幅広い分野からなる特許文書を対象とした場合においても、語や表現の違いを吸収した対比を実現することができる。例えば、オシロスコープの別表現として、波形観測装置や電圧波形測定器等が得られる。これらの語は、特許以外の文書で使われることは少ないが、特許文書の対比においては有効である。

以上のように、本願と引例候補文献の並列スペクトルバー表示と、シソーラス辞書の適用により、文書データの一覧性を高め、かつ、語句や表現の違いを吸収したスペクトルバー表示が可能になるため、対比のための作業時間を短縮することができる。

2.3 適用事例の紹介

図3は、シソーラス辞書を適用する前後でのスペクトルバー表示と、得られる対応結果の差異を表した事例である。(図3)左側は、本願の発明特定事項と、発明特定事項を特徴付けるキーワードを強調表示により示す。(図3)右上側には、指定したキーワードをそのままスペクトルバー表示した結果を点線で、(図3)

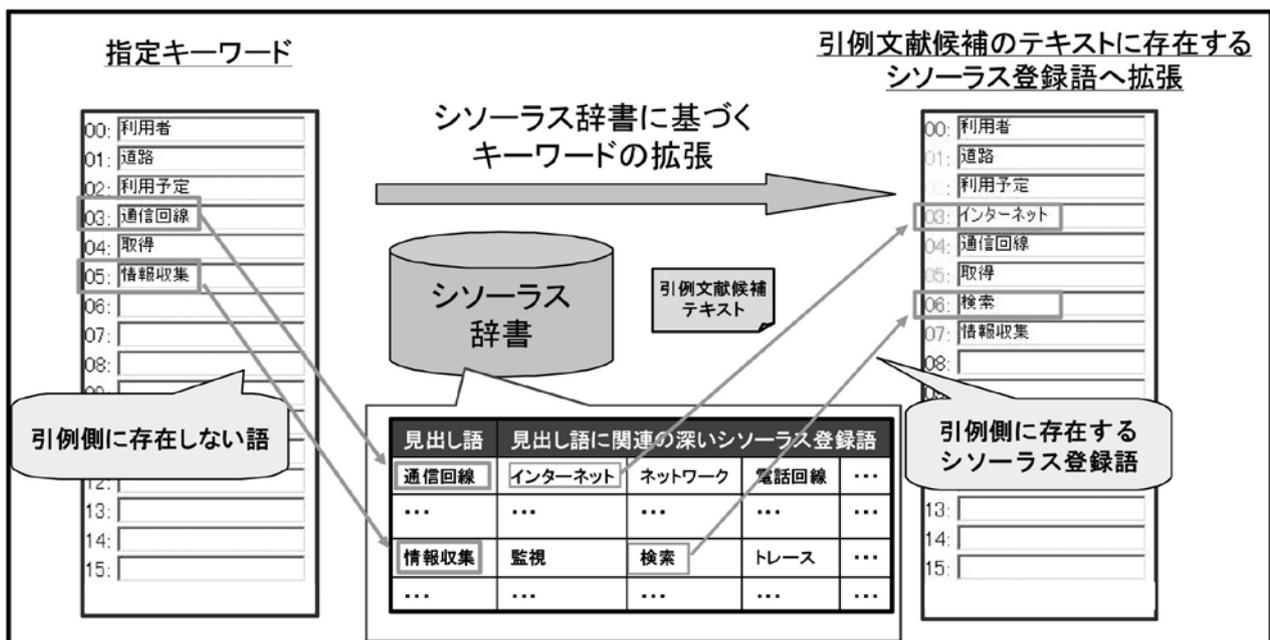


図2 シソーラス辞書に基づく、ユーザが指定したキーワードの拡張イメージ

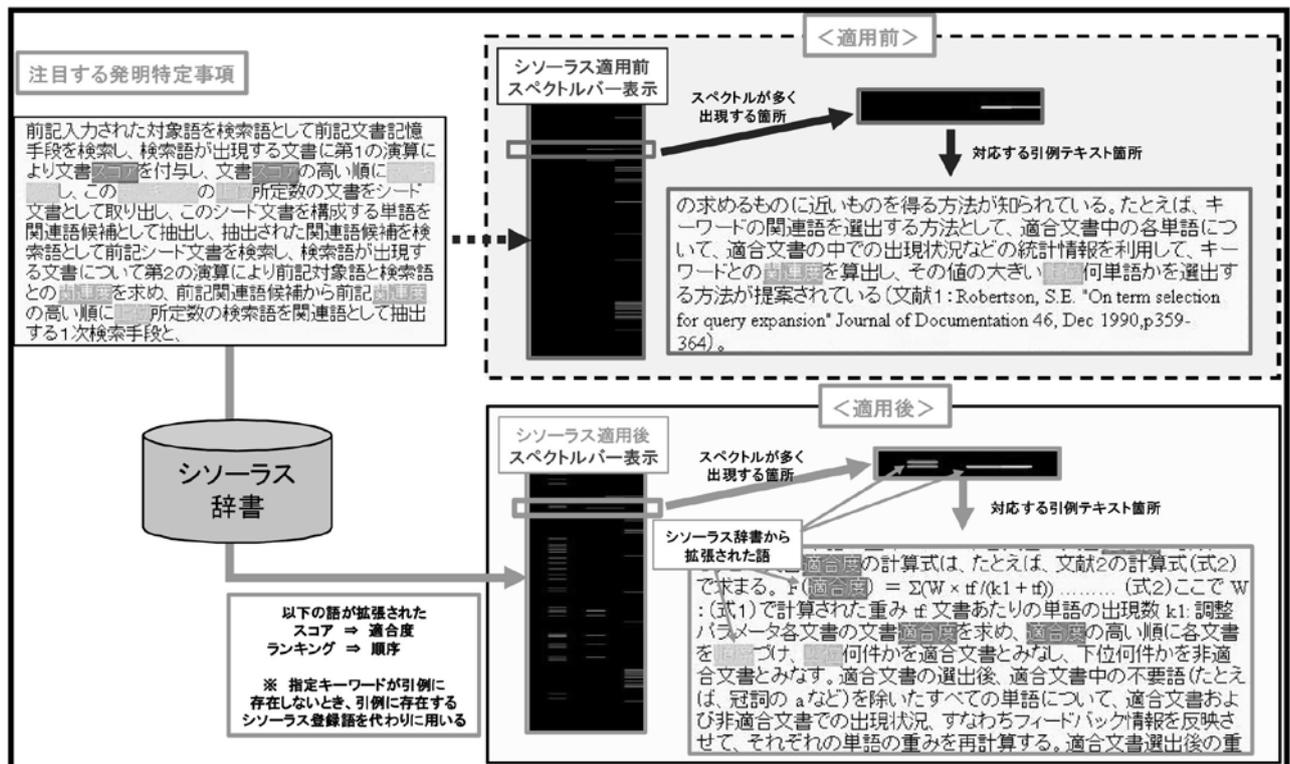


図3 シソーラス辞書適用前後のスペクトルバー表示と得られる対応箇所

右下側には、キーワードをシソーラス辞書に基づき拡張したときのスペクトルバー表示結果を実線で示す。

シソーラス辞書を用いて、引例側に存在する語に限定してキーワードを拡張すると、スペクトルバー表示部のスペクトルが増加する。これは、特定の文字列内にユーザが指定したキーワードと、その拡張語が存在することを示す。この情報を活用することで、入力キーワードに関連する記載部分の発見が容易になる。実際にスペクトルバー表示結果に基づき得られた対応箇所を(図3)の(a)と(b)に示す。シソーラス辞書適用前の結果(図3(a))は、発明特定事項に関連する内容が記載されているが、類似先願であることを特定する記述としては不十分である。一方、シソーラス辞書を用いてキーワード拡張(「スコア」に対して「適合度」、「ランキング」に対して「順序」)することで、本願の発明特定事項を詳細に説明する部分(図3(b))を取得できた。

3 おわりに

本稿では、審査における特許調査の効率化支援の一つとして、本願と引例文献候補の対比を行う特許対比支援システムを紹介した。本システムは、本願と引例文献候補の並列スペクトルバー表示により情報の一覧性を高め、膨大かつ特許特有な用語を含むシソーラス辞書を対比に適用することで対比精度を高めることを可能とした。

現状で吸収可能な語句や表現の違いは、語と語、語とフレーズのレベルである。しかし、文献によっては、ある単語で表される概念が、文章や段落というように、より広い範囲で表現されることがある。表現範囲の差を吸収した対比を実現する方法については今後の課題である。