

生成AI時代における特許検索方法の改良提案

—生成 AI を活用した多言語キーワードの取得と査読、および補助ツールの作成—

Proposal for improving patent search methods in the era of generative AI



アジア特許情報研究会

西尾 潤

2015 年からアジア特許情報研究会で中国特許検索および知財情報処理について研究。現在は科学技術振興機構に勤務する傍ら（一社）日本知的財産協会及び企業主催の中国特許検索セミナー講師、（一社）情報科学技術協会（INFOSTA）主催 情報プロフェッショナルシンポジウム（INFOPRO）実行委員を務める。アジア特許情報研究会 note: https://note.com/sasiapi_

✉ jnishio10@sasiapi.org

1 はじめに

昨年（2025 年）は特に自然言語をベースとした生成 AI が社会基盤として認められた年といえよう。知財業界においては 2015 年ころから機械翻訳での活用が始まり、2019 年ころから知財業務の特定タスクに特化した AI が開発・販売されてきた^[1]。そして、生成 AI の出現により自然言語分野で汎用性を有しながら、これら特定タスクを代行するまでの存在に進化してきたといえる。加えて、AI によるハルシネーション対策として引用先を示すこと、機密情報漏洩の問題に対しクラウドのテナント内で完結するシステムの構築やローカル LLM（大規模言語モデル）の使用が現実的になったことなどが社会基盤として認められうる要因の一つと思われる。

現時点では、汎用生成 AI のみで特許データベースにアクセスして高度な検索を行うことは不可能なため、特許検索実務の全自動化は困難である。しかし、LLM の技術理解力を活用した査読方法が提案されている^[2]。

本項では特許検索実務をすべて生成 AI に任せられるのではなく、従来の特許検索実務の中に生成 AI を取り入れる方法を実験データと考察により提案する。

2 提案手法

本項では次の手法を提案する。

- 1) 生成 AI を用いて多言語キーワードを取得する。
- 2) 従来の検索方法に準じて検索集合を作成する。

- 3) 検索集合 α に対し、発明の名称・要約・請求項に特定のキーワードをかけたサブセットを複数（5～8 個程度）作成し、サブセットにいくつ属するか（以後、構成要件充足数という）で分類する。
- 4) 構成要件充足数が特定条件を満たすものを残し、それ以外を振るい落とす。
- 5) 非日本語・非英語文献の場合、2 か国語検索で見つかった集合と、どちらか片方の集合に分け、上記 3、4 の操作を行う。
- 6) 5 までの操作で残った文献の一覧表示で、キーワードごとに設定した着色を行う「アノテーション」機能を用い、文献一覧表示で文字列に付された色情報を参考にノイズ落としする。この機能を生成 AI に作成させた Excel のマクロで代替できる。
- 7) 特定のキーワードが明細書のどこに位置するかを示

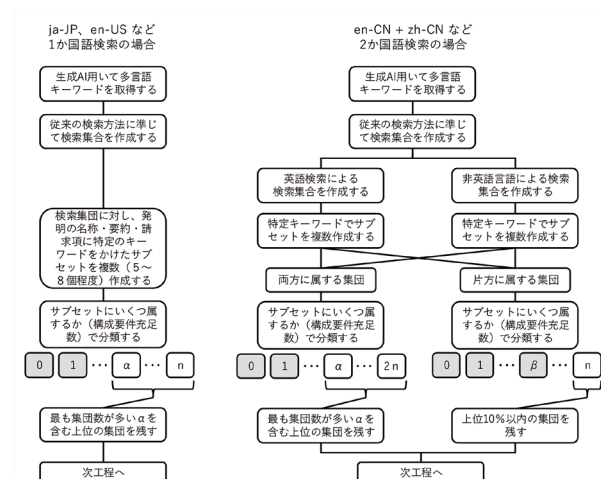


図1 提案手法の1～5を模式図にしたもの

す「ミニマップ」機能を用い、ノイズ落としする。生成 AI に作成させたプログラムで代替できる。

8) 複数の文献を取り込むことができる生成 AI を用いて、特定の構成要件の記載根拠を出力させる。

従来の手法では、キーワードや特許分類の掛け合わせを種々変えて小さな検索集合を複数作り、その和集合を検索集合とすることでノイズを減らし網羅性を高める方法がとられてきた^[3]。

本提案手法は、ある文献に検索式における構成要件を多く記載されていれば検索意図に近い文献であるし、1個も記載されていなければ全く関係がないという仮説から、従来法の基本を守ったうえで、統計的手法で査読順を付ける手法である。本提案では、図 1 に示す操作の範囲で、文献の文字を読むことなくスクリーニングを行うことが特徴である。

以下、実際の手法と適用順が異なるが、実験結果を示しつつ提案手法の詳細を解説する。

実験では得られた全ての文献に対し、あらかじめ査読してリチウムイオン電池・微生物・金属抽出・Mn, Co, Ni の 4 要件を含む文献を類似 (○)、そのうち 3 要件を含む文献を参考 (△)、残りを非類似 (×) として 3 分類して検証に用いた。

本稿では分類モデルの性能評価に用いられる ROC 曲線の考え方を応用し、横軸を査読数、縦軸を○△公報の発見率 (最大 1.0) を表す GAIN 曲線を示す。曲線の肩が左上にあるほど分類性能が高い (査読効率が良い) ことを示す。なお、青線はそれより左の文献を査読し、それより右の文献を除外する境界を意味する。

3 実験に使用した検索式

実験ではリチウムイオン電池を回収してその正極材料に含まれる Mn, Co, Ni 金属を取り出す技術において乳酸菌や酢酸菌などの微生物を適用するという内容の発明があったときに、この発明の特許性を否定する先行技術文献調査を行うことを想定した検索式を作成した。

検索式は、キーワード 8 行、IPC8 行、論理式 50 行からなる。各論理式は従来通り発明の構成要件のうち 2～4 つを用い、さらに必要に応じて他の要件をかけて小集合を作り、最後にその和集合を取る方法で作成した。

通常の検索実務では一例としてヒット数が 60 件を

図 2 日本語キーワードによる検索式の一例

超える論理式 (白色で表示) を除外するなどして件数を減らす。除外した式に○公報が存在する可能性が排除できない。そこで本実験では FTO (Freedom to Operate) 調査のように網羅性が必要なケースを想定して、本来除外すべき式も用いた。その結果、日本語検索集合は [66] 式の 173 件から [67] 式の 504 件に増加した。

次に、中国特許文献の検索は、日本語検索式の日本語キーワードを生成 AI で取得した英語のキーワードに変えてグローバル検索し、得られた中国特許文献の集合と、他国の文献の中国ファミリーの集合と合わせて英語検索による中国特許文献 (en-CN) の集合を得た。一方、日本語検索式のキーワードを生成 AI で取得した中国語キーワードに変えて中国特許文献を検索し、中国語検索による中国特許文献の集合 (zh-CN) を得た。

4 統計処理によるスクリーニング

4.1. サブセットを用いた統計的手法

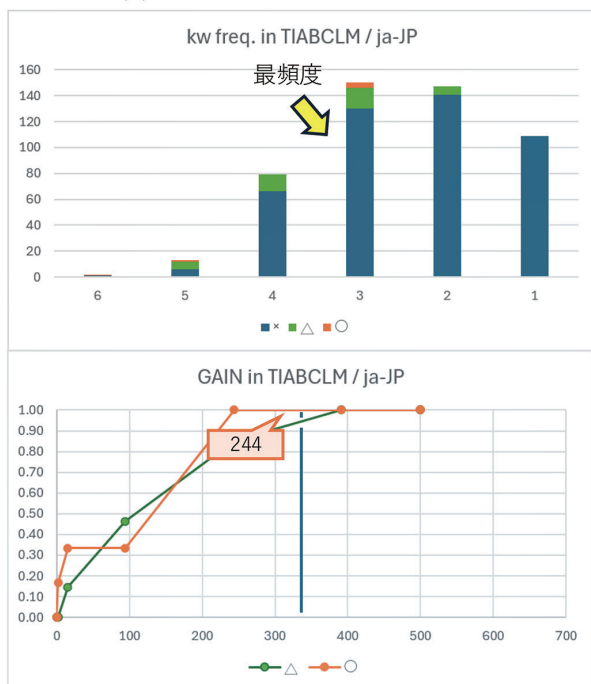
日本語で日本特許を検索した集合 (ja-JP) について、それぞれ検索時のキーワードと類義関係の 8 種類のキーワード群を絞り込み検索した 8 個のサブセットを作り、構成要件充足数を集計しグラフにした。なお○を橙色、△を緑色、×を青緑色で塗り分けている。

8 個のキーワード群は、①リチウムイオン電池、②正極材料、③微生物、④菌種、⑤有機酸、⑥浸出、⑦Mn, Ni, Co、⑧回収、に関する類義語の集合とした。

また、キーワード絞り込み対象は、発明の名称・要約・請求項と明細書全文の 2 条件で比較した。

図 3(a) では要件充足数最多の 8 から文献数が最も文

(a) TI+AB+CLMで絞り込んだ時



(b) 全文で絞り込んだ時

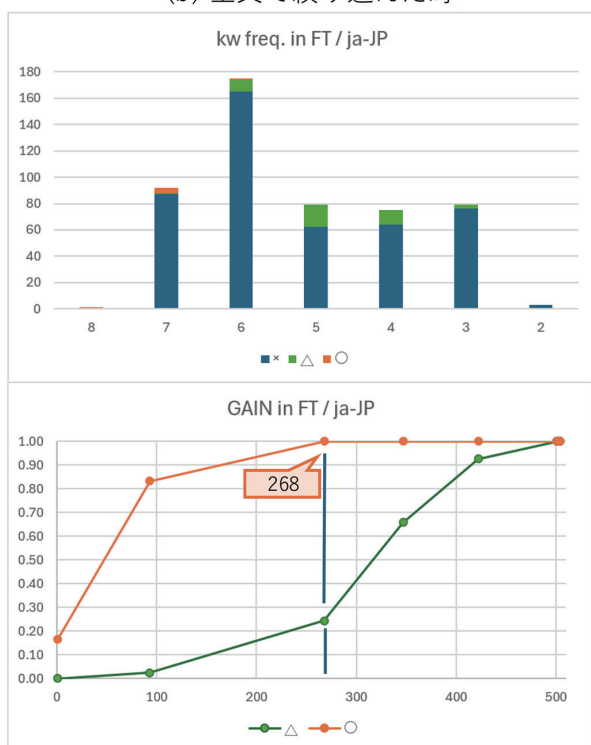


図3 ja-JP 構成要件充足数分布とGAIN曲線。(a) 発明の名称・要約・請求項キーワード絞り込み結果を元にしたデータ、(b) 明細書全文キーワード絞り込み結果もとにしたデータ

献数が多いランク（以下、最頻値ランク）3まで漸増し、その間に○の文献がすべて含まれる。一方、図3(b)では、最頻値が6で、その間に○の文献がすべて含まれる。

構成要件充足数分布からわかるように、最頻値より下のランクに属する文献を捨てる操作で○をすべて残すことができる。一方、△文献について(a)条件の方が(b)

条件より残りやすいことがわかった。

実際の調査の場面では○△×の判定はされていないから、検索集合に発明の名称・要約・請求項のキーワードで絞り込み検索した複数のサブセットを作り、最頻値ランク以上のサブセットの文献を読むという運用方針がよいと思われる。その結果、本提案手法の3、4が定められた。なお、最頻値ランクより後ろのサブセットに○文献が含まれている可能性について9節で救済を試みる。

中国特許を英語で検索した時、検索集合を(a)の条件で絞り込んだ場合、最頻値ランク以上のサブセットに○文献がすべて入ったが、検索集合を(b)の条件で絞り込んだ場合、最頻値ランクより下位に○文献が存在した。中国特許を中国語で検索した時も同様であった。

この条件(a)について考察した。(a)の絞り込み条件は要約のみでよいという考えもあるが、要約を持たない文献や、外国語文献を中心に、「明細書に記載の発明。」や「本願はPCT出願で要約は国際段階で公開。」など、構成要件が記載されていない要約を持つ文献もあることから、(a)の絞り込み条件を発明の名称・要約・請求項とするのが良いと思われる。

このときの統計処理は図4に示すように、A列に検索集合の出願番号、B～I列にサブセットの出願番号を転記し、J列にCOUNTIF関数を用いた構成要件充足数の式を記入して計算し、最後にJ列を基準に降順ソートする。公報番号でなく出願番号を扱うのは、同一出願のA公報とB公報などの重複を避けるためである。

SUM	A	B	C	D	E	F	G	H	I	J
	出願番号	出願番号_1	出願番号_2	出願番号_3	出願番号_4	出願番号_5	出願番号_6	出願番号_7	出願番号_8	構成要件充足数
	特許2025-581016	特許2025-561374	特許2025-552358	特許2025-552358	特許2025-561374	特許2025-561374	特許2025-561374	特許2025-561374	特許2025-561374	8
	特許2025-552358	特許2025-552358	特許2025-552358	特許2025-552358	特許2025-552358	特許2025-552358	特許2025-552358	特許2025-552358	特許2025-552358	7
	特許2025-540395	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	7
	特許2025-530331	特許2025-540395	特許2025-525064	特許2025-530331	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	特許2025-547866	7
	特許2025-112570	特許2025-530331	特許2025-517483	特許2025-7365	特許2025-540395	特許2025-540395	特許2025-540395	特許2025-540395	特許2025-540395	7
	特許2025-575065	特許2025-525064	特許2025-534202	特許2025-517385	特許2025-530331	特許2025-530331	特許2025-530331	特許2025-530331	特許2025-530331	7
	特許2024-554870	特許2025-517483	特許2025-527006	特許2025-506761	特許2024-575056	特許2025-525064	特許2025-525064	特許2025-525064	特許2025-525064	7
	特許2025-112570	特許2024-527006	特許2025-112570	特許2024-556139	特許2025-112570	特許2025-517483	特許2025-517483	特許2025-517483	特許2025-517483	7
	特許2024-533084	特許2025-85844	特許2024-528334	特許2025-85844	特許2024-554870	特許2025-85844	特許2025-517385	特許2025-517385	特許2025-517385	7

図4 構成要件充足数を計算する様子

4.2. 論理式への落とし込み

4.1節では、8つのサブセットを作り、そのサブセットに含まれる数が多いほうから査読していくことを提案したが、わざわざワークシート上で計算するのは面倒である。これを検索式で処理する方法について述べる。

図5に示すように、検索集合が式[67]、絞り込みキーワードが式[68]～[75]とすると、構成要件充足数8の集合を求める論理式[76]は、[67]*([68]*[69]*...*[74]*[75])で表される。次に要件充足数7の集合を

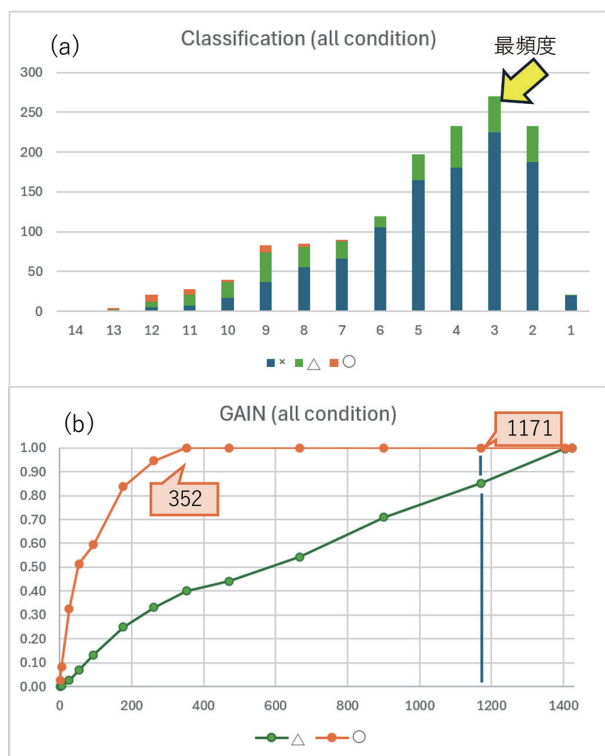


図7 実験①で得られた集合の(a) 構成要件充足数ランキングと、(b) GAIN 曲線

集合 (B) と (C) をまとめて集計した。それぞれ構成要件充足数ランキングと GAIN 曲線を図 8 に示す。

実験①の場合、352 件の査読数で○文献すべてを取得できるが、最頻値のサブセットまで読むルールでは検索母集団 1425 件のうち 1171 件読む必要がある。

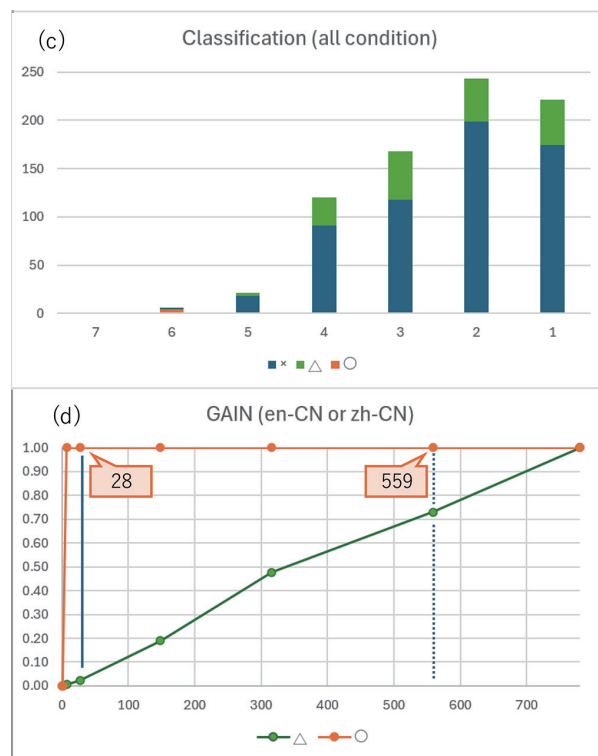


図8 実験②で得られた集合 (A) の (a) 構成要件充足数ランキングと (b) GAIN 曲線、集合 (B)(C) の (c) 構成要件充足数ランキングと (d) GAIN 曲線

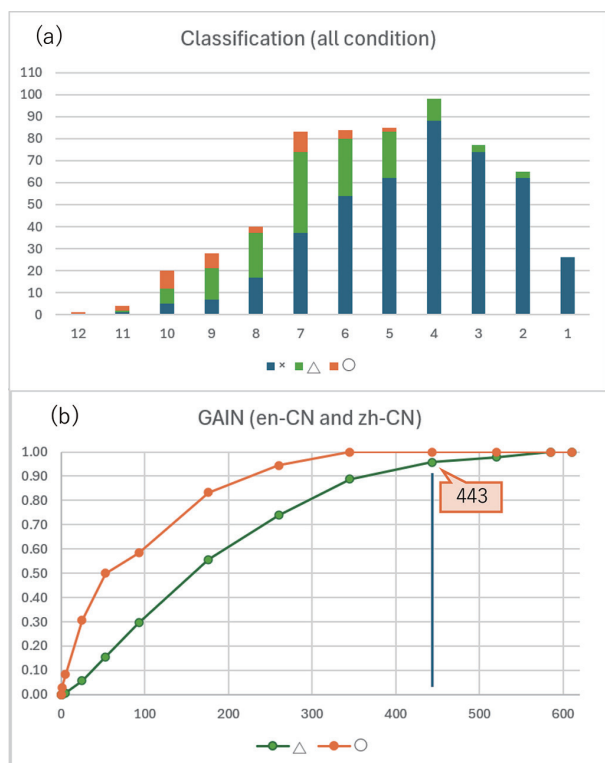
一方、実験②の集合 A では最頻値のサブセットまで読むルールに従って 443 件読むと○文献すべてを取得できることが分かった。

実験②の集合 B について、7 件読めば○文献すべてを取得できるが、最頻値のサブセットまで読むルールに従うと、559 件読む必要がある。実験②については集合 (A) に漏れた文献のうち、必要な構成要件を満たす文献は集合 (B) で上位に集約されることが図 8(c) から読み取れる。そこで、集合 (B) は構成要件上位のサブセットから読んでいき、総数 10%以内のサブセットに含まれる文献を読むとよいと思われる。これより検索母集団 1425 件の文献を読まずに 471 件に絞り込むことができた。

5 キーワードの豊富化

通常、検索式において同一カテゴリにおけるキーワード同士は OR 検索（並列）として扱われる。そのため、検索キーワードが増えると検索集合の文献数が増加する。検索集合の中に一定割合でノイズが含まれることを考えると、検索キーワードが増えると査読数が増えて検索効率が低下するよう思える。

ところが、本手法においてキーワードを豊富化させた



ことにより査読数が減った事例に遭遇したのでこれを報告する。

なお、この事象は追加した検索キーワードが当該技術分野において使用される用語であるときに生じる。

[illegible]

図9 図2の検索式でキーワードの数を増やしたもの。赤字が追加したキーワード

図 9 は図 2 で示した検索式に、赤色の文字で示される技術用語を追加したことを示す図である。追加した技術用語は (1) 生成 AI にリチウムイオン電池の回収技術を検索させ、(2) 英語・中国語キーワードを出力させ、(3) 英語・中国語キーワードを入力して日本語の類義語を出力させ、(4) 技術分野に照らして妥当であるか、他の用語の一部にならないかを基準に検索集合に悪影響を及ぼさないとされる用語を人為的に選択して追加したものである。

検索集合に悪影響を及ぼす事例としてはリチウムイオン電池の略号である LIB で検索をかけるとスペルに「lib」を含む菌種が多数ヒットする。また、英語検索の例で元素名の Ni や Co の文字は技術用語でない単語に多く含まれる。これらの用語を含んだ検索式を立てると検索意図を反映しないものになる。そこでキーワードの人為的選択が必要になる。

図 2 はキーワード豊富化前の検索式で検索集合は 504 件、図 9 は豊富化後の検索式で検索集合は 701 件になった。

ここで提案手法に従い構成要件充足数が少ない部分集合をカットすると、査読数が 391 件から 358 件に減

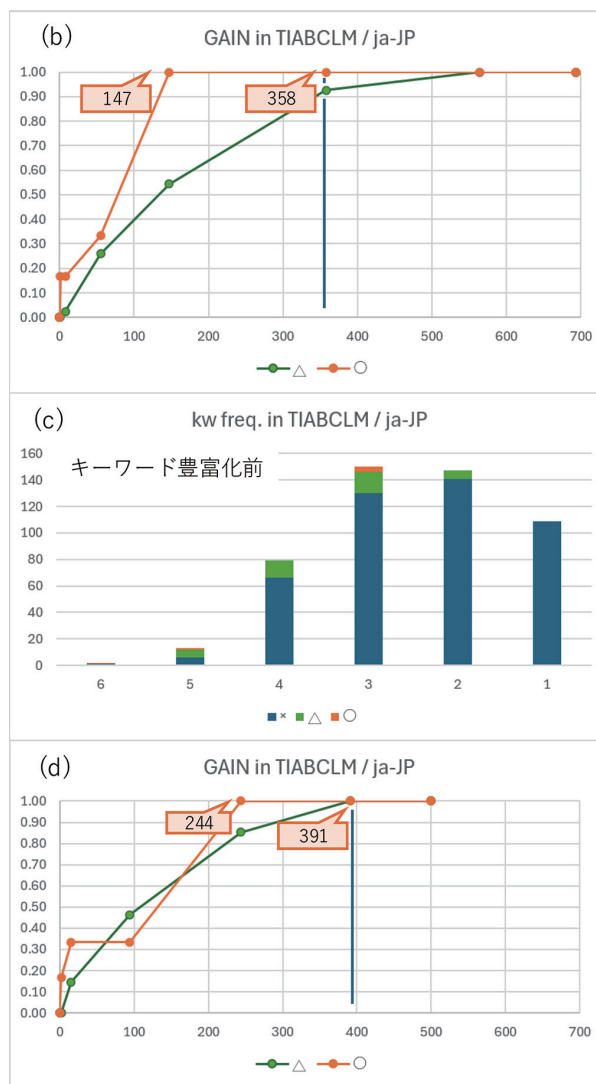
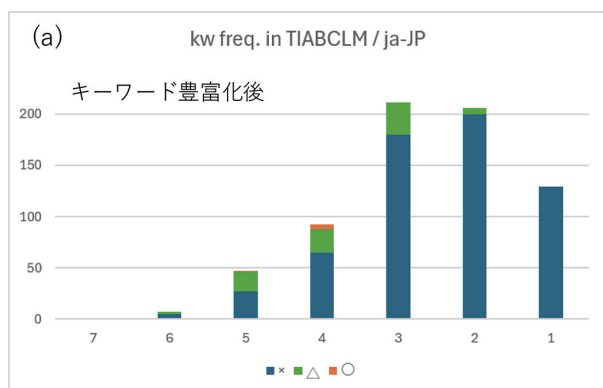


図 10 キーワード豊富化後の (a) 要件充足数グラフと (b) GAIN。
キーワード豊富化前の (c) 要件充足数グラフと (d) GAIN

少した。また、○文献を 100%補足できるまでに必要な査読数が、244 件から 147 件へと減少した。

図 10(a) は (c) に対して○文献を示す橙色の表示が、構成要件充足数の大きい側に移動して、充足数 4 までに集約されている。この現象は追加したキーワードが発明の名称・要約・請求項に含まれ、構成要件充足数が増加したことを示している。本手法では構成要件充足数が最頻値より少ない集合をスクリーニングでカットするので、上記の通り検索集合数が大きくなったにもかかわらず査読数が少なくなったと考えられる。

ここで、仮に追加したキーワードが当該技術分野で用いない用語であったとする。そうするとサブセットの絞り込みにヒットしないので、○文献であっても構成要件充足数が増加せず、構成要件充足数の最頻値も少ないほうに移動する結果、査読する文献数が増加すると考えられる。そのため、キーワードを豊富化する操作では生成

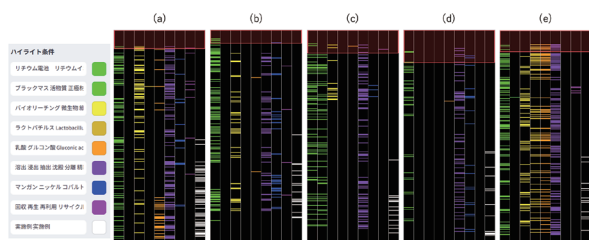


図 13 ミニマップ表示ツールに 5 つの公報を読み込ませ表示した状態。ここでは (c) ~ (e) を振るい落とす

最後まで全体にわたって分布している。公報 (c) は微生物を表す黄色が公報の前半にしかないため、先行技術として微生物が記載されていて、本発明に直接関係ないと思われる。公報 (d) は微生物を用いない金属回収方法と思われる。公報 (e) は Mn・Ni・Co を表す青がないため、そのほかの金属種を微生物浸出すると思われる（実際には鉄を抽出している）。

なお、白線で表すように「実施例」の文字列の位置を参照すると、発明特定事項が実施例に記載されている可能性が高いことが示される。

ただし英語公報の場合、実施例を示す文字列として“example”を指定すると、本文中の“for example”（たとえば）という多数の表記をミニマップが表示して正確な実施例の場所を判別できないことがある。同様に“Black mass”の“Black”など一般名詞として用いられる単語や、ミニマップを表示するための HTML のタグがノイズになることがある。

商用の特許データベースには図 11 のテキストアノテーションツールや図 13 のミニマップの機能が搭載されていることが多いので、積極的に使用することを推奨したい。またテキストアノテーションツールはクラウド型^[4]や OSS^[5]としても提供されており、これを用いると非特許文献のスクリーニングにも適用できる利点がある。

図 13 のツールは、「pdf ファイルを読み込み、画面左に本文テキストを表示させ、画面右にミニマップを表示させるプログラムを python で作る」こと、「ミニマップはテキストと色の条件に従ってその位置を線で表示させる」こと、「テキスト表示はテキストと色の条件に従って該当する文字列の背景色を設定すること」などを生成 AI に指示して作成したものである。

pdf ファイルを読み込ませるツールを使う場合、pdf ファイルにテキストが埋め込まれている必要がある。画像だけで構成され埋め込みテキストがない pdf ファイ

ルを使うときは、これを一度 Microsoft Word で読み込ませてから pdf で保存すると、レイアウトは崩れるがテキストが埋め込まれたファイルが得られる。

8 生成 AI を用いた査読

ここからは二次スクリーニングになる。明細書本文にどのような内容が記載されているか読んでいくのであるが、ここで生成 AI が威力を発揮する。

この目的に合う生成 AI を選定するときに重要な事項は、①多言語に対応していること、②技術内容を正しく理解し説明できること、③コンテキスト長（入出力可能な文字の長さ）が長いこと、④取り込んだファイルをもとに回答を生成し、外部に検索に行かないこと、が必要で、この目的に合うのは検索拡張生成（RAG）やクラウドストレージが備える AI、NotebookLM（Google）を代表とするソースに基づく回答をする AI である。最近の高性能な生成 AI ではハルシネーション対策として回答の根拠を示すようになった。RAG や NotebookLM では、データを（チャンクと呼ばれる）適当な長さに切断し、その中から質問文に近いチャンクを検索してその文脈に近い回答を出力するが、NotebookLM はチャンクの長さが長いと、文脈の意図をくみ取った回答が得られやすい。

ここでは NotebookLM を使った査読方法を記述する。

まず、同時に取り込めるファイルは 50 までの制限がある。そこで事前に十分スクリーニングを行って文献数を減らしておくことが望まれる。

複数のファイルを読み込ませると、全体に共通する内容を抽出して、それに対応する英語タイトルが付与され、日本語による要約文が出力される。

プロンプトとしては以下のものが有用である。

- ・（特定のソースを指定して）この文献にはどのようなことが記載されているか教えて
- ・（複数のソースを指定して）これらの文献の中に（特定の構成要件）が記載されている文献を探して
- ・列名を構成要件 1、構成要件 2、… 構成要件 n、行のタイトルを文献名にして、それぞれの列に関する記載があればそれを書き出し、なければ「記載なし」とした表を作成して（対比表の作成）
- ・各ソースについてそれぞれ（特定の構成要件）がど

各ソースについて、最近を適用する前にどのような処理を行うかを表にまとめて

提供された各ソースに基づき、細菌（微生物）を適用（バイオリッチング等）する前に行われる処理（前処理）の内容を以下の表にまとめました。

ソース	対象物	細菌適用の前に行われる処理（前処理）
O95_0007_0011.pdf	酸性鉱山廃水由来等の細菌株	細菌を培養し、DNAを単離および精製する。その後、16S-23S rDNAスペーサー領域をPCR増幅し、RFLP解析や配列決定の準備を行う ①。
115_172.pdf	黄銅鉱コンцентレート	表面に吸着したキサンテン酸を除去する ④。B.E.T.法により比表面積を測定する ④。
CN107083484A	廃旧リチウム電池	放電、粉碎、水との混合、NaOHによるpH調整（13.0～13.5）、60～70℃での攪拌反応、濾過、塩酸によるpH調整（5.0～6.0）を経て水酸化アルミニウム沈殿を回収し、濾液を浸出に使用する ⑤。 NaOHおよびNaCl溶液に12時間浸漬して放電 ⑦。解体して正極材

図 14 NotebookLM を用いて特定の構成要件が記載されている文献を出力させた様子（誤字があっても当該技術分野を基に正しく理解している）

のような効果を生じているか教えて

このようにして外国文献も非特許文献も日本特許文献と同列で扱うことができ、効率的に査読が行える。

9 生成 AI による落穂拾い

本提案手法のポイントは公報を読まずにスクリーニングすることである。そのため、スクリーニングでふるい落とされた公報にも〇文献があることは否定できない。

そこで、生成 AI を用いて以下の実験を行った。

zh-CN 文献を Mn, Co, Ni でスクリーニングしたサブセット（〇文献を含む）から 1000 件の公報リストをダウンロードして、得られた Excel ファイルを汎用生成 AI（Gemini）にアップロードし、発明の名称と要約から①リチウムイオン電池、②微生物 / バイオ処理、③ Mn, Co, Ni の回収がすべて記載されている出願番号を出力させたところ、4 件の出力があったが、そのうち 2 件は実在しない出願番号で、残り 2 件も全く違う分野で、完全に間違っていた。

次に、NotebookLM を用いて同様に行った。ただし NotebookLM は .xlsx を受け付けられないため .pdf に変換してファイルをアップロードした。そうすると 3 件出力され、すべてが正解であった。このようにして正解公報を抽出することができた。

NotebookLM の技術理解力が高いことから、検索集合の段階で NotebookLM を使用してスクリーニングすることが視野に入るが、NotebookLM は 1 ソースあたり 200 MB または 50 万語のファイルの制限がある

ため、出願番号 + 発明の名称 + 要約 + 請求項のデータを用いると検索集合全件を処理するのは難しいと思われる。さらにソースとして pdf ファイルを準備する手間がかかることと生成 AI が毎回異なる回答を返すことを考えると、本提案手法の一次スクリーニングでふるい落とされた公報リストを入力して、スクリーニング漏れを救済するために用いることが良いと思われる。

G列（発明の名称）とT列（要約）を確認し、ご指定いただいた3つの条件（①リチウムイオン電池、②微生物/バイオ処理、③Mn, Co, Niの回収）が揃っているものを抽出した結果、以下の出願番号（C列）が該当しました。

抽出結果：該当する出願番号

条件をすべて満たす、または強く示唆する出願は以下の通りです。

出願番号（C列）	判定理由（キーワードの抽出）
202180026214.X	名称に「バイオ浸出」、要約に「リチウムイオン電池」「ニッケル、コバルト、マンガンの回収」の記載あり。
202180025791.X	名称に「バイオ浸出」、要約に「リチウムイオン電池の正極材料」「Ni, Co, Mnの浸出」の記載あり。
201980064227.4	要約に「微生物（アシディチオバチルス属等）を用いた廃リチウム電池からの有価金属（Ni, Co, Mn）回収」の詳述あり。
202211603590.2	名称・要約に「微生物法」「リチウムイオン電池」「ニッケル、コバルト、マンガンのリサイクル」の記載あり。

図 15 NotebookLM を用いて一次スクリーニングでふるい落とした文献を再調査する様子

10 生成 AI を用いたキーワードの取得

5 節「キーワードの豊富化」、6 節「テキストアノテーションによるスクリーニング」、7 節「ミニマップによるスクリーニング」で見た通り、検索やスクリーニングの場面で多言語キーワードを豊富に取得することが好ましい。

ここでは一般的な生成 AI を用いて、当該分野の用語を出力させる。しかし 9 節でみたように汎用生成 AI は特許検索が機能不全であるため、研究論文を抽出させてその分野のキーワードを出力させる方法が良い。

プロンプトとしては以下のものが有用である。

- ・（無料論文検索サイトの URL）で、リチウムイオン電池を回収して、微生物を用いて Mn, Ni, Co を抽出したことが記載された論文を 10 報出力して
- ・この分野で正極材料についてどのような技術用語があるか、論文から原文のまま忠実に抜き出して出力して（これにより英語の技術用語が抽出される）
- ・Black mass, Active material, Cathode material, Positive electrode, Ternary material の日本語による言い換え（これにより日本語の技術用語が抽出される）
- ・ブラックマス、活物質、正極材、三元系の中国語に

リチウムイオン電池を回収して、金属を分離する方法が記載された論文10報の、それぞれの金属同士の分離方法を検索

細菌（微生物）を検索する際の英語キーワードは何？

細菌（微生物）を用いた金属分離・回収の研究を英語で検索する際、より専門的で精度の高い論文を見つけるためのキーワードを整理しました。

これらを組み合わせることで、Google ScholarやScopus、各ジャーナルのサイトで目的の「Article（論文）」にたどり着きやすくなります。

1. 「細菌・微生物」に関するキーワード

単に「Bacteria」と打つよりも、研究分野に合わせた用語を使うのがコツです。

- Microorganisms / Microbes: 微生物（全般）
- Biohydrometallurgy: バイオ湿式精錬（細菌を利用した金属回収の学問分野名）
- Metal-reducing bacteria: 金属還元細菌（Coなどを還元して分離する細菌）

微生物+微小生物+細菌+菌体混合物+混合培養系+微生物群の中国語による言い換え

これまで整理してきたバイオリサイクル（生物湿式製錬）の文脈で、中国語の学術論文や技術レポートに登場する「生き物」たちの呼び分けを整理しました。

中国語では、単一の菌を指すのか、それとも協力し合う「集団」を指すのかで言葉が明確に使い分けられます。

1. 「微生物」そのものを指す言葉

日本語	中国語（簡体字）	ピンイン	ニュアンス
微生物	微生物	Wēishēngwù	最も一般的な総称。
微小生物	微小生物	Wēixiǎo	微生物より少し広い範囲（微小な多細胞生物含）

図 16 生成 AI を用いてリチウムイオン電池を回収して金属を取り出す日英中技術用語を取り出す様子（抜粋）

よる言い換え（城市采矿という新用語が発掘される）

- ・ 城市采矿の日本語と英語による言い換え

生成 AI は技術説明を付けるので、それを読んで検索意図に合致する技術用語を収集することができる。

11 おわりに

本稿は検索集合をキーワードで絞り込むことにより、少ない文献査読数で効率よく特許調査を行う方法を提案するものである。

生成 AI に特許調査を全部任せるには、① 検索式を正確に処理し返却するツールの整備、② 技術分野を検索し適切なキーワードを取得するエージェントの整備、コンテキスト長が長く 1000 を超える文献を処理できる RAG の実現などが必要であるが、それらの実現はさほど非現実的ではないと思われる。

本稿は 2026 年度の「アジア特許情報研究会」および「中国特許検索」セミナーの改定に伴うワーキングの一環として報告するものである。なお、本稿の内容は個人の見解に基づくものであり、所属組織を代表するものではない。

参考文献

- [1] 野崎篤志, 情報の科学と技術 (2018) 68、7、316-325.
- [2] 安藤俊幸, Japio YEAR BOOK (2024) 218-229
- [3] 角淵由英, パテント (2022) 75、5、3-15
- [4] FastLabel(<https://fastlabel.ai/>), LabelBox(<https://labelbox.com/>) など
- [5] brat(<https://brat.nlplab.org/>), Label Studio(<https://labelstud.io/>) など