

## 特許調査におけるAI等の活用に関する研究

情報検索委員会  
第2小委員会\*

**抄 録** 近年、AI等を用いた特許調査ツールが注目されており、会員企業各社においても業務効率化の観点からこれらの導入を検討する場面は増えてきているものと推測される。一方でそれらのツールについて、具体的にどのような場面で、どの程度利用できるのかといった実力の把握や、より効果的に活用するための方法の模索を一企業のみで行うのは、リソース面から容易ではなく情報が不足している。そこで本研究では、対象業務として特許無効化のための先行文献調査、および技術動向調査のための分類付与を想定し、各種特許調査ツールの実力把握のための比較検証を行い、その結果に基づき目的に応じた使用ツール選定とその組合せの考え方、また機械学習を併用した調査の効率化等について検討した。なお、調査対象のツールはAIの搭載を謳ったものに限定せず、従来から提供されていたツールも含め、各対象業務に適用しうる機能を用いて検討した。

### 目 次

1. はじめに
2. 検討概要
  2. 1 検討対象業務
  2. 2 評価指標
3. 無効調査
  3. 1 対象案件
  3. 2 対象ツール
  3. 3 評価手順
  3. 4 評価結果と考察
4. 動向調査
  4. 1 対象案件
  4. 2 対象ツール
  4. 3 評価手順
  4. 4 評価結果
5. おわりに

### 1. はじめに

知財部員にとって特許調査は、自社事業分野における技術動向の把握、出願前の先行文献の検討、事業の実施に向けたクリアランス調査な

ど、どれを取っても重要な業務であるのは言うまでもない。しかし、毎年各国で新たな出願がなされている以上、成長期や成熟期等の事業段階による程度の差はあるとしても基本的に調査対象となる特許は増え続けており、さらに昨今では非特許情報も考慮した分析への期待も高まっていることから、調査担当者への負担も年々増加してきている。

このため、特許調査の効率化がますます求められるようになってきており、その方法の一つとして、AI等を利用した特許調査ツールの活用が注目されている。会員企業各社においても、これらのツールの導入や活用に向け、具体的にどのような場面で、どの程度利用できるのかといった実力や、より効果的な利用方法等の情報に関心が寄せられているものと推測されるが、一企業のみでこれらを検討するのは、工数面や費用面から容易ではない。

\* 2019年度 The Second Subcommittee, Information Search Committee

表1 特許庁でのAI活用への取組の例<sup>1)</sup>

	業務名	対象	業務内容・課題	ソリューション案
1	電話等の質問対応	全課室	担当課室の特定、問合せ対応 (関係課室からの、のべ6業務を統合)	導入可能性あり
2	紙出願の電子化 登録商標の使用の確認	特許	紙出願された資料を、OCR(光学的文字認識)によ り読み取り、登録商標の使用の確認	導入可能性あり
5	品質監査 (作成書類の誤記確認)	特許	審査官の作成書類の記載の形式的な瑕疵の確認	導入可能性あり
6	特許分類付与	特許	出願された発明に対し、発明の技術的事項に基づき、 適切な分類を付与	一部導入可能性あり
7	先行技術調査	特許	出願された発明に関連する先行技術文献を、データ ベース等を活用して抽出	一部導入可能性あり
8	発明の内容理解・認定	特許	出願された発明の内容を理解し、特許保護を求めて いる技術的思想の技術的範囲を認定	現時点では導入困難
9	特許登録可否の判断	特許	出願された発明と過去の発明を対比し、特許を付与 すべきかを判断して、判断の具体的理由を示す拒絶 理由通知等を作成	現時点では導入困難
			出願意匠に対し、日本意匠分類・ロカルイ分類を付	

そこで本研究では会員企業の参考となる情報を提供すべく、各種ツールについて検討したため、以下に報告する。

なお本研究は、2019年度情報検索委員会第2小委員会第1ワーキンググループの青山裕樹(小委員長, ポリプラスチックス), 南宅崇人(ダイキン), 村松慎吾(帝人), 飯村信(第一三共), 久我範夫(マレリ), 峯尾泰(富士フィルム知財情報リサーチ), 吉永泰子(三井化学)によるものである。

## 2. 検討概要

### 2.1 検討対象業務

冒頭で述べた通り、特許調査には様々な目的があるが、本研究においては対象業務として、特許無効化のための先行文献調査(以下、「無効調査」とも言う)および技術動向調査のための分類付与(以下、「動向調査」とも言う)を想定して検討を行った。詳細は後述するが、これらはともにベンチマークとなる客観的なデータを入手しやすく、かつ実務的に導入へのハードルが比較的低い<sup>2)</sup>と思われることから検討対象に選定した。なお、これらの業務に関しては

表1のように特許庁でのAI活用への取組においても「一部導入可能性あり」とされており、加えて表2の通り近年ベンダー各社からのAI搭載ツールのリリースが進んでいる点からも期待が高まっている領域と考える。

表2 AIを搭載したツールの例<sup>3)</sup>

ツール名	ベンダー
スクリーニング効率化・レイティング	
KIBIT Patent Explorer	FRONTEO
Xlpat	Xlpat Labs
Patent Noise Filter	アイ・アール・ディー
Deskbee	アイ・ピー・ファイン
Patentfield	IP Nexus
amplified ai	amplified ai
Innovation Q Plus	IP.com
クラスタリング・分類展開	
Patent Mining eXpress Text Mining Studio Visual Mining Studio	NTT データ数理システム
Xlpat	Xlpat Labs
Patent Predictive Analyst	アイ・アール・ディー
Nomolytics®	アナリティクスデザイン ラボ
Shareresearch 分析オプション <sup>3)</sup>	日立製作所、ニッセイコム
Derwent Data Analyzer (Record Auto-Classifer)	クラリベイト・アナリティ クス
新規性・進歩性判断	
IP Samurai	ゴールドアイビー

ただし、これも詳細は後述するが、本研究では検討対象を「AI搭載」を積極的に謳ったツールに限定せず、従来から提供されていたツールも含め、各対象業務に適用しうる機能の活用を検討した。

## 2. 2 評価指標

本研究では前述の対象業務について、「人手による調査」の結果に対する「各種ツールによる調査」の結果との比較を行った。ここで「人手による調査」というのは、例えば検索式の構築や抽出特許の判別を人間が検討・判断して行ったものを指し、「各種ツールによる調査」というのは、AI利用有無に拘らず、各種ツールに搭載された、調査の自動化や効率化を支援する機能を利用し、人間の作業を極力排除して行ったものを指す。

この評価指標としては、過去から文書検索の精度評価に用いられてきた<sup>4)</sup>「再現率（漏れの少なさ）」および「適合率（ノイズの少なさ）」の考え方を利用した。ただし、通常用いられる

再現率と適合率の定義（図1）で計算すると、本研究の事例では適合率が極端に低くなり比較が困難となりうるため、これを調整する意味で、以下適合率の算出方法を変更（発見不可案件の除外<sup>5)</sup> および実質調査件数での分母設定<sup>6)</sup>）する場合があるが、その際は便宜上「補正適合率」と称する。また、次章以降では再現率や適合率（補正適合率）以外の指標を用いている箇所もあるが、それらは必要に応じ適宜説明する。

## 3. 無効調査

### 3. 1 対象案件

無効調査における「人手による調査」として、過去の日本の無効審判にて、先行する日本特許出願を証拠に、新規性および／または進歩性の理由により無効化された特許計50件（以下、「対象特許」とも言う）を選定<sup>7)</sup>し、この証拠に用いられた先行文献計119件（表3参照）を検出できるか、各種ツールによる調査を実施した。

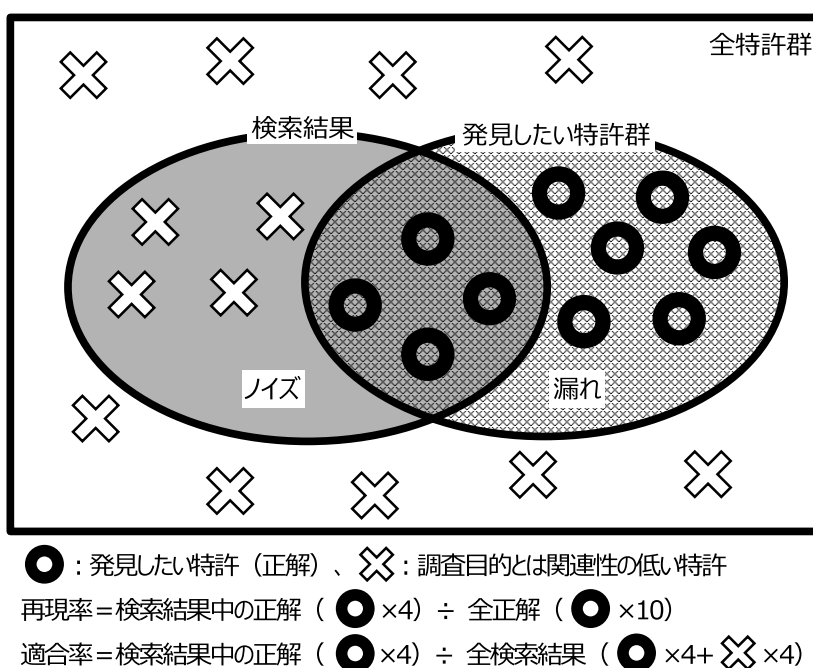


図1 再現率と適合率の定義のイメージ

表3 対象案件の概略内訳<sup>8)</sup>

分野	無効理由	対象特許件数	先行文献件数
機械	新規性	7	10
	進歩性	16	44
化学	新規性	10	18
	進歩性	4	15
電気	新規性	3	5
	進歩性	10	27
合計	新規性	20	33
	進歩性	30	86

### 3.2 対象ツール

各種ツールによる調査では、表4および表5に示したツールを使用した。このうち表4のツールは、指定した検索条件に該当する特許を出力する、いわゆる「特許検索ツール」であり、表5のツールは母集団中の特許に対し特定の条件を基にスコアを付与する、いわゆる「レーティングツール<sup>9)</sup>」である。

なお、表4の特許検索ツールは、主に本研究

を実施した委員が所属している各企業にて以前から導入していたツールで、特段「AI搭載」を特徴としている訳ではないものが多いが、いずれも「ある特許の番号を入力することにより、その特許の技術に類似する内容の特許を検索する機能（以下、「番号類似検索機能」とも言う）」を有する、またはさらに「ある技術の概要を文章で入力することにより、その技術に類似する内容の特許を検索する機能（以下、「テキスト類似検索機能<sup>10)</sup>」とも言い、上記の番号類似検索機能と合わせて「類似検索機能」とも言う）」を有するものを選定した。ここで表4には各ツールの番号類似検索機能の概略および出力可能な類似特許検索結果の抽出件数上限を合わせて示した。

また表5のレーティングツールは、いずれも「AI搭載」を特徴としており、母集団および教師データとなる特許を与えることで、母集団に含まれる特許について、教師データに対する類似度のスコアを付与する機能を有している。

表4 無効調査（類似検索）の検討に使用した特許検索ツール

ツール	番号類似検索機能概略	抽出件数上限
A	名称・要約・請求項からキーワードを自動抽出し検索	20,000
B	請求項中のキーワードのベクトルと頻度を比較	10,000
C	機械学習（詳細不明）	無し
D	統計的にキーワードを自動抽出し検索	300
E	特許分類の頻度と引用関係で検索	2,000
F	自動抽出したキーワードと引用関係で検索	無し
G	特許分類と引用関係で検索	無し
H	キーワードを自動抽出・重み付けしスコア付与	20,000
I	キーワードを自動抽出・重み付けし検索	500

表5 無効調査（母集団中の特許の並べ替え）の検討に使用したレーティングツール

ツール	教師データ数	母集団件数上限
$\alpha$	1件で可（複数設定することも可）	100,000
$\beta$	1件で可（複数設定することも可）	無し
$\gamma$	推奨：100件程度～	10,000
$\delta$	推奨：100件程度～	データのファイルサイズ10MB以下

### 3. 3 評価手順

本研究では、まず表4のツールにより表3の対象特許50件の類似検索<sup>11)</sup>を行い、「無効審判で証拠として採用された先行文献のうちどの程度を抽出できるか(再現率)」および「どの程度の件数の母集団を確認することで目的の先行文献に到達できるか(補正適合率)」という観点から、人手による調査に対する各種ツールによる調査の結果を評価した。さらに番号類似検索により得られた類似特許群を母集団とし、これを表5のツールにより類似度スコアの高い順に並べ替えること(以下、「並べ替え機能」とも言う)で補正適合率向上の可能性を検討した<sup>12)</sup>。

なお、上記各工程については過去にも幾つか検討事例<sup>13), 14)</sup>が報告されているが、それらはあくまで類似検索機能や並べ替え機能を個々に評価したものであった。これに対し本研究では、各機能を併用することで、検索式の構築や抽出特許の精査等の人手による作業を極力排除した効率的な無効調査の可能性を検討した。

### 3. 4 評価結果と考察

#### (1) 類似検索機能の評価結果

##### 1) 番号類似検索機能の評価結果

各ツールの番号類似検索機能による無効調査の結果を「横軸：再現率、縦軸：補正適合率」として図2に示す。すなわち図中の右上方向に向かうほど再現率と補正適合率がいずれも高く有利な結果となることを意味する。

なお、図中でマーカーの脇に付したA~Iは表4中の各ツールの記号に対応し、マーカーの形状は類似検索機能の機序(表4の「番号類似検索機能概略」の仕様が近いものは同じ形状)、マーカーの塗り潰しは抽出件数上限(白抜きは上限無し、ハッチングは1万件から2万件、黒塗りは1万件未満)をそれぞれ表す。

この結果からは、一部例外はあるものの全体

の傾向としては、当然ながら「抽出件数上限が多いツールでは漏れが少ないがノイズは多く(右下領域)、抽出件数上限が少ないツールでは漏れが多いがノイズは少ない(左上領域)」という再現率と補正適合率のトレードオフの関係が確認された。

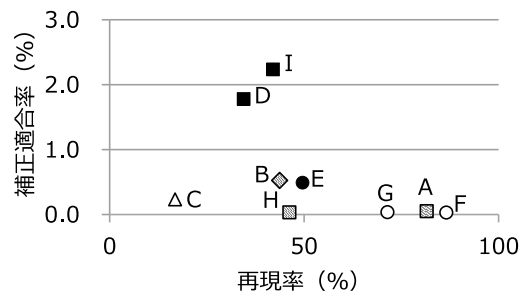


図2 番号類似検索機能での無効調査結果

また、各ツールによる再現率と補正適合率を対象特許の分野別に算出してプロットした結果を図3に示す。

図中のF値とは「 $2 \times (\text{再現率} \times \text{補正適合率}) \div (\text{再現率} + \text{補正適合率})$ 」により求められる「調和平均」を指し、再現率と補正適合率との総合的な評価指標として用いられる値である。これが「機械分野：2.6>化学分野：1.1>電気分野：0.6」の序列であることから、機械分野では化学分野や電気分野に比べ、再現率や補正適合率が相対的に高い傾向にあることが分かる。

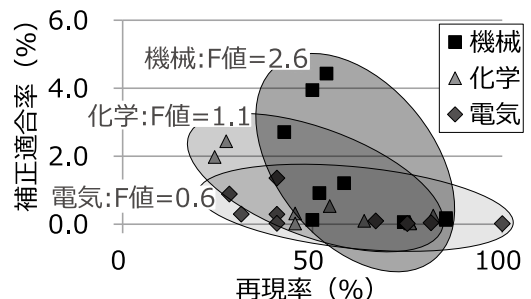


図3 番号類似検索結果の分野による傾向

この傾向は図中に楕円状ハッチングで示した



各分野のプロット領域が上記と同じ順で右上寄りに分布していることから見て取れる。

このように機械分野に対して化学分野や電気分野が不利となる理由としては、これらの分野では、請求項や明細書に化学式や数式を含む場合が多いことが影響しているものと考えられた。

## 2) テキスト類似検索機能の評価結果

番号類似検索において、ツールE～Gが高い再現率を示した点に関して、これらのツールの類似検索機能は引用関係も考慮していることが影響している可能性が懸念された<sup>15)</sup>。

そこで引用関係の影響を除外するため、表4のツールのうちテキスト類似検索も行えるものを用いて、対象特許の請求項と要約を検索条件として入力しテキスト類似検索を行った。

結果は番号類似検索と同様に、再現率と補正適合率のトレードオフが確認された(図4)。

なお、図4中のアルファベットおよびマーカーの形状と塗り潰しの意味は図2と同様である。

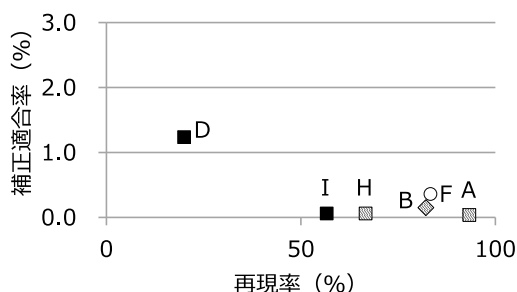


図4 テキスト類似検索機能での無効調査結果

## 3) 結果の考察

上記にて述べた通り、各種ツールによる無効調査においては、基本的に再現率と補正適合率がトレードオフの関係にあり、残念ながら再現率も補正適合率も高い(図の右上領域にプロットされる)ツールは確認されなかった。すなわち各種ツールによる調査では、人手による調査に比べ検索式の構築にかかる工数がある程度省

略することはできるとしても、その検索によって得られた母集団の確認においては、人手による調査と同様に「目的とする先行文献を漏らさず発見するためには多量の母集団を確認する必要がある、逆に工数を抑えるために母集団の件数を減らせば目的とする先行文献を発見できない可能性が高くなる」ということになる。

ただし、これはあくまで表4の各種ツールの抽出件数上限に基づいて得られた母集団の中の特許を全件確認するという前提での話であり、実務的には、限られた時間の中で調査を行うに際し「この程度の工数で目的とする先行文献に到達したい」という、ある程度の母集団件数の期待値があるものと思われる<sup>16)</sup>。そこで、今回の先行文献が母集団中の何件目で発見されたか、つまり類似検索により得られた母集団において各先行文献が類似度順で何位に含まれていたかの分布を図5として整理した。

なお、図5の横軸のアルファベットA～Iは表4の各種ツールを、その下線は抽出件数上限(二重下線は上限無し、一重下線は1万件から2万件、下線無しは1万件未満)を、縦軸の数値範囲は番号類似検索機能により得られた各母集団中において表3の先行文献が含まれていた類似度順の範囲を、マトリクス上の各円の大きさと円内の数値は当該類似度順範囲内で発見された先行文献の件数(すなわち円内の数値を縦軸方向に合計した値はA～Iいずれも119件)を、各円の色は類似検索機能の機序(図2と同様に分類し仕様が近いものは同じ色で表示)を、それぞれ示す。

図5から、前述の母集団件数の期待値として仮に500件<sup>17)</sup>を想定した場合、図中に破線枠で示す通り、(一部例外はあるが)類似検索機能の機序等によらず概ね先行文献119件中の50件程度が当該範囲に含まれており、各種ツールで同等の結果を得られることが確認された。

しかしながら、先行文献119件のうち50件(再

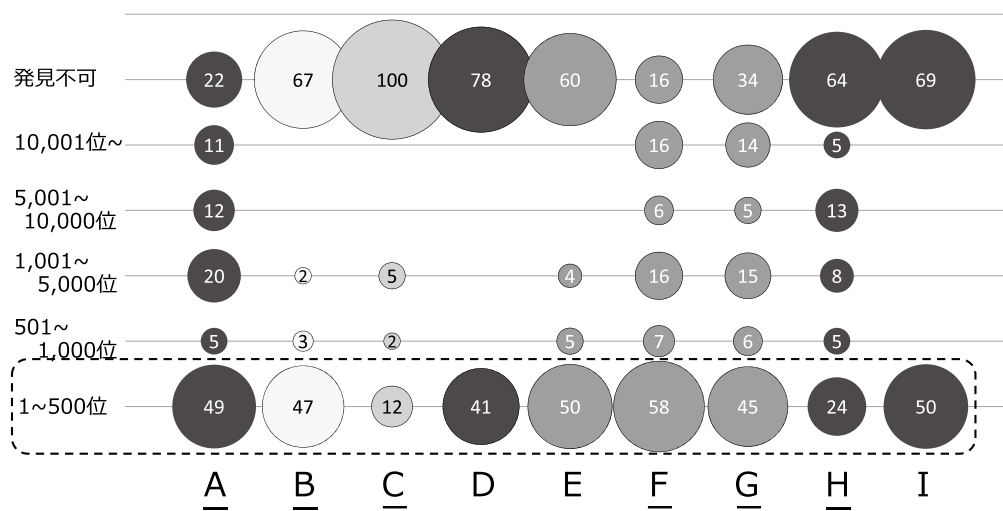


図5 番号類似検索機能での無効調査により発見された先行文献の類似度順分布

現率として約40%) というのは、約60%の先行文献が発見できないということであり、無効調査としては心許ない。そこで無効調査の目的として先行文献を確実に発見できることが最重要とすれば、まず再現率を高くするために図2で右寄りにプロットされたツール(図5で「発見不可」が少ないツール)を選定することが必要となる。ただし、再現率と補正適合率のトレードオフからそのようなツールではノイズが多くなる。例えば図5にてツールA, F, Gのような「発見不可」の件数が少ないツールを見ると、目的とする先行文献を発見するまでに数千件から一万件以上の特許を確認する必要があるケースも一定数存在することが分かる。

そこで、このような高再現率のツールによる無効調査について補正適合率の向上を検討した。

## (2) 並べ替え機能の評価結果

高再現率のツールによる無効調査結果の補正適合率を向上する手法として、3章2節~3節で述べた通り、表5に挙げた「AI搭載」を特徴とするレーティングツールを用い、表3の対象特許を教師データとして高再現率の母集団中の特許に類似度スコアを付与し、それによる並

べ替えを行い、目的とする先行文献の類似度順を上昇させることでの補正適合率の向上を試みた。

ここで上記の通り、本評価では教師データとして各対象特許1件のみを用いたが、使用したツールには表5の $\gamma$ や $\delta$ のようにある程度教師データの数が多いことが推奨されているものも含まれている。この推奨教師データ数の違いは、搭載するAIの違い、さらに言えば各ツールの想定用途(開発思想)の違いにより生じるものであり、その意味では $\gamma$ や $\delta$ は本評価に適したツールとは言えない可能性はあるが、そのような違いが結果にどう表れるかを見る目的で評価対象とした。

なお本評価では、各種レーティングツールを横並びで比較できるよう、それぞれのツールの仕様(処理可能な母集団件数上限等)を踏まえ、対象案件を表3の対象特許50件から、さらに10件(その先行文献28件)に絞って実施した。

またツールや評価条件は表6に示す通り、1)レーティングツールによる差、2)特許検索ツールによる差、3)教師データによる差、4)解析フィールドによる差の4点から検討した。

表6 並べ替え機能の評価ツール<sup>18)</sup>と評価条件<sup>19)</sup>

レーティングツール	特許検索ツール (母集団作成用の特許検索ツール)	教師データ (教師特許からのデータ抽出範囲)	解析フィールド (解析対象特許のデータ参照範囲)	
1) レーティングツールによる差 $\alpha$ $\beta$ $\gamma$ $\delta$	2) 特許検索ツールによる差 C F G	全請求項	類似度Top10段落	
		自動		
		請求項1		
		請求項1+効果		
	3) 教師データによる差	G	全請求項	類似度1位の段落 請求項1
		H	全請求項	
	4) 解析フィールドによる差	H	全請求項	類似度Top10段落
			全請求項	名称+要約+全請求項
			名称+要約+全請求項	名称+要約+全請求項
			名称+要約+全請求項	名称+要約+全請求項

1) レーティングツールによる差

特許検索ツールHにより得た母集団を用いて、レーティングツール $\alpha \sim \delta$ で並べ替えを行った結果、図6に示すように各ツールとも並べ替え前に比べ補正適合率の上昇が確認された。なお、ツール $\alpha$ 、 $\beta$ がツール $\gamma$ 、 $\delta$ より高補正適合率となった理由は先述の通り、推奨教師データ数が異なるためであり、教師データ1件のみにて行った本評価として、ツールの仕様通りの妥当な結果が出たものと考えられる<sup>20)</sup>。

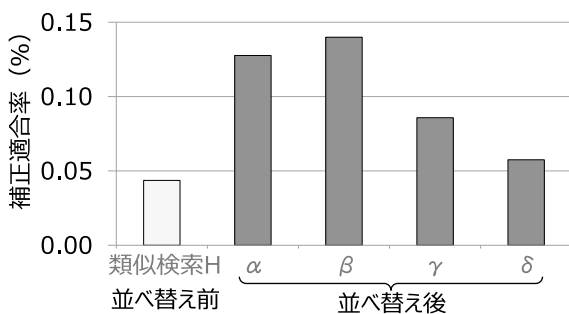


図6 レーティングツール別での並べ替え結果

また、これを図5と同様の類似度順の分布としたものを図7に示す。各ツールでの並べ替え

により、元は一万件以上の確認が必要であった先行文献の件数が減少し、特にツール $\alpha$ 、 $\beta$ はツール $\gamma$ 、 $\delta$ よりも類似度順が高い件数が多く、効率的に目的の先行文献に到達できうることがここからも見て取れた。

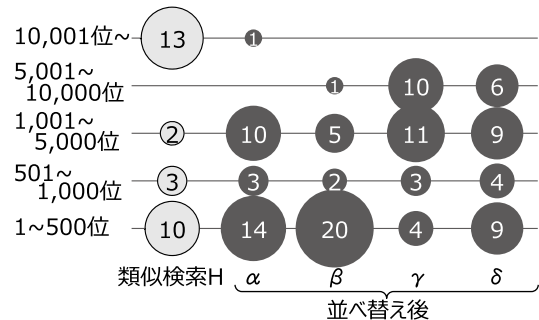


図7 並べ替え後の先行文献の類似度順分布

2) 特許検索ツールによる差

レーティングツール $\alpha$ を用いて、異なる特許検索ツールで作成した母集団の並べ替えを行い、特許検索ツールによる差を評価した<sup>21)</sup>。

前述の通り、用いる特許検索ツールによって類似検索結果の再現率と補正適合率は異なるが、図8の通り、いずれの類似検索結果を母集団に



用いた場合も並べ替え後の補正適合率の向上が確認された（ただし母集団内の順番を並べ替えたのみであるため再現率は変化していない）。

なお、図8中のアルファベットは母集団作成に用いた表4の各特許検索ツールを、マーカーの形状は特許検索ツールの類似検索機能の機序（図2と同様に分類し仕様が近いものは同じ形状で表示）を、マーカーの塗り潰しは並べ替えの前後（白抜き：並べ替え前、黒塗り：並べ替え後）を、矢印は各ツールでの並べ替え前後における補正適合率の変化を、それぞれ意味する。

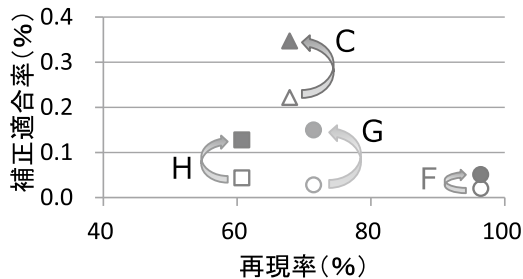


図8 母集団別での並べ替え結果

### 3) 教師データによる差

特許検索ツールFにより得た母集団を用いて、レイティングツールaによる並べ替えを行うにあたり、教師データとして対象特許のどの範囲を使用するかによる差を評価した<sup>22)</sup>。

図9は横軸をツールaでの並べ替えにおける各教師データとして、図5と同様に類似度順の分布を示したものである。なお、教師データのうち「自動」はツールaが備える機能により、教師データとする範囲を自動的に選択する（ただし請求項1は必須とする）ものである。

この結果からは、教師データとしてどの範囲を用いた場合も、類似度順5,001位以上の範囲に該当する件数が並べ替え前より減少しており、並べ替え後の集合では、より早い段階で発見が可能な先行文献が増加（すなわち補正適合率が向上）したことが窺える。中でも教師データの範囲に、「請求項1+効果」を選択した場

合は特に有利と見られる結果であった。

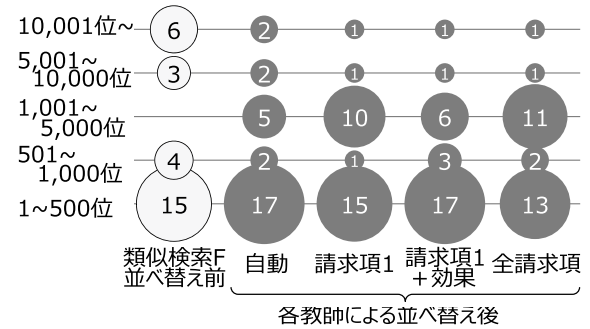


図9 教師データ別での並べ替え結果

### 4) 解析フィールドによる差

特許検索ツールGにより得た母集団を用いて、レイティングツールaによる並べ替えを行うにあたり、母集団中の各特許についてどの範囲を解析対象（解析フィールド）として教師データとの類似度を判定させるかによる差を評価した。なお、この評価はツールaに搭載された機能により各請求項および明細書中の各段落に付与された類似度スコアをもとに、図10の横軸の各フィールド（これもツールaの機能として用意されていた選択肢から設定）での類似度スコアを集計し、その順位で母集団中の特許を並べ替えて行ったものである<sup>23)</sup>。

図10に示す通り、いずれの解析フィールドにおいても、並べ替え前に対する補正適合率の向上が見られたが、類似度スコア順の上位10段落のスコア合計値に基づく並べ替え（類似度Top10段落）を用いた場合に特に補正適合率が向上する結果が得られていた。

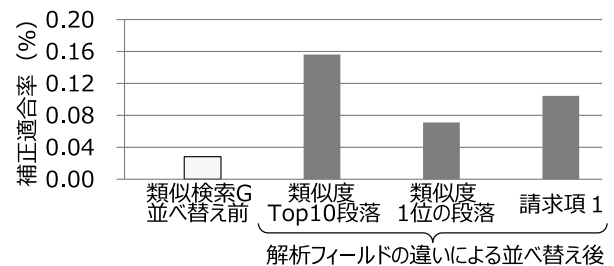


図10 解析フィールド別での並べ替え結果

### 5) 結果の考察

以上の通り、並べ替え機能の活用による無効調査の補正適合率向上の可能性が示唆された。ただし表7の特許5042638のように、並べ替えにより類似度順が低下するケースも確認された。

また並べ替え条件の選択においても、例えば教師データについては、対象特許の「請求項1+効果」を選択した場合が特に有利であったと述べたが、表7を見ると、必ずしも全ての対象特許において「請求項1+効果」を教師データとした場合に最も類似度順が高くなるという訳では無く、いずれの教師データでも類似度順が最も高くなる場合がありうるということが確認できる。

これを考慮すると、実務上においては1種類の教師データで並べ替えた集合だけを確認するより、複数種の教師データで並べ替えた複数の集合を得ておいて、それらを上位から一定件数(例えば100件)ずつ横断的に確認した方が、より早く目的の先行文献に到達しうる、つまり先述の「母集団件数の期待値」の範囲内で無効調査を完了できうるものと考えられる。とは言え、教師データの種類が過度に多いと、当然ながら各ツールの解析処理にかかる時間も長くなる上、内容を確認しなければならない集合自体の数も増えることになり、却ってより多くの調査期間

を必要とすることになりかねない。このため、教師データの種類としては例えば2~3種程度に留めることが推奨される<sup>24)</sup>。

これらを踏まえ、各種ツールの選定や活用の方法について以下に整理する。

### (3) 小 括

前項までの検討に基づき、各種ツールの活用による無効調査の効率化手法として「高再現率の母集団の補正適合率向上」を検討する上での主なポイントを表8に挙げた。

これを纏めると、「抽出件数上限が多い特許検索ツールの類似検索機能により得た高再現率の母集団を、少量の教師データでの解析が可能なレイティングツールを用いて並べ替えを行うことで、対象特許に対する類似度スコア順付与の高精度化を図り、さらにこの操作を複数種の教師データについて行うことで得られた複数の集合を、特定件数ごとに横断的に確認する」と表すことができる。

この手法により、目的とする先行文献の発見に要する工数の低減が可能になると考えるため、無効調査の効率化に関する本研究の一つの成果としてこれを提案する。

表7 教師データの違いによる並べ替え結果における類似度順の例

対象特許	先行文献	母集団：特許検索ツールFでの番号類似検索により得られた集合				
		並べ替え前 (ツールF)での 類似度順	教師データによるツールαでの並べ替え後の類似度順			
			自動	請求項1	請求項1 +効果	全請求項
特許5042638	特表2005-511808	21	36	37	225	225
特許4543046	特開2003-103996	43,079	46	34	1,416	2,851
	特開2004-276416	43,075	5,180	4,430	2,956	4,748
特許3887652	特開2005-127608	804	2	2	44	1
特許4136457	特開2003-225153	503	90	59	59	356
特許4015734	特開平11-192064	733	88	108	41	105

表8 各種ツールによる無効調査の効率化のポイント

工程	検討項目	ポイント
対象特許からの類似検索による母集団作成	特許検索ツール選定	類似検索機能の抽出件数上限が多い（または無い）ツールを使用
対象特許に対する類似度スコア順での並べ替え	レイティングツール選定	教師データ量が少量での解析を得意とするツールを使用
	教師データ設定	請求項，要約，効果等，およびそれらの組合せによる複数のパターンを使用
並べ替え後の集合の確認	集合の確認順序	複数の教師により得た複数の集合を類似度スコア順で上位から一定件数ごとに横断しつつ確認

## 4. 動向調査

### 4.1 対象案件

次いでもう一つの対象業務に挙げた動向調査に関し、「人手による調査」として日本特許庁が実施した「特許出願技術動向調査」から、平成28年度の「ファインバブル技術<sup>25)</sup>」を選定し、この特許文献リストより抽出した日本出願3,381件を対象として、日本特許庁が付与した表9の大分類1～7への分類を、各種ツールによってどの程度再現できるかを検討した。

なお、大分類1～7の該当公報件数の合計が3,381件を超えていることから分かるように、対象特許1件について複数の分類が付与されているものも存在する。

表9 大分類1～7の概略

No.	大分類のカテゴリ名称	人手分類による該当公報件数
1	ファインバブル発生技術	1,261件
2	ファインバブル計測技術	303件
3	ファインバブルの種類材料	1,334件
4	ファインバブル技術の課題目的	1,806件
5	ファインバブル技術の解決手段	1,364件
6	ファインバブル応用技術	2,664件
7	ファインバブル利用の目的効果	2,408件

### 4.2 対象ツール

ここでは、前章の無効調査にて表5に挙げたレイティングツール $\gamma$ と $\delta$ を用いた。この2ツールでは教師データの件数をある程度多くすることが推奨されている。このため教師データに対象特許1件のみを使用していた前章の無効調査では、ツール $a$ や $\beta$ に比べ補正適合率の向上が小さく不利と見られる結果となっていたが、本章の動向調査では後述する通り教師データが100件と比較的多いため、良好な解析結果が期待されるものである。

### 4.3 評価手順

教師データは、日本特許庁が人手により付与した大分類をもとに、ある大分類に対し当該大分類に該当する公報（正解教師）と該当しない公報（ノイズ教師）とを50件ずつ無作為に抽出した計100件として、これらが大分類1～7のそれぞれについて用意した<sup>26)</sup>。

上記教師データをもとに、各大分類について解析対象特許3,281件（教師データ100件以外）がそれぞれどの大分類に該当するか、ツール $\gamma$ および $\delta$ により判定を行った。この該否判定の閾値としてはサーチ確率（当該大分類に該当する確率として各ツールにて算出された値）が50%以上のものを「該当」とした。また評価指標には再現率と適合率を用いた。ここで再現率は

「ある大分類について、人手による判定で当該大分類を付与された公報（正解）のうちどれだけがツールによる判定でも当該大分類を付与されていたか」を、適合率は「ある大分類について、ツールによる判定で当該大分類を付与された公報のうちどれだけが正解であったか<sup>27)</sup>」をそれぞれ意味する。

なお、今回のような複数項目（7項目）への分類付与を行う場合、「解析対象の各公報について各大分類への該否判定（該当か非該当かの二値分類）を、7回繰り返し行う」か、「各公報について大分類1～7のいずれか1分類への類別判定（7項目への多値分類）を一度に行う」という2通りの方法が考えられる。この点、先述の通り日本特許庁の人手による分類では、1件の対象特許に対し複数の分類が付与されたものが存在していたことから上記二値分類に近いものと思われるが、本研究では多値分類についても評価した。

また教師データとして「特定の大分類のみが付与された公報を50件選択する」のか、「特定の大分類に加え他の大分類も付与された公報を50件選択する」のかの差による影響も評価した（以下では便宜的に、前者の特定の大分類のみの公報からなる教師データを「単一教師」、後者の特定の大分類以外の大分類も含んだ公報からなる教師データを「複合教師」とも言う）。

#### 4. 4 評価結果

##### 1) ツールによる差

ツール $\gamma$ と $\delta$ を用いて、単一教師<sup>28)</sup>により、各大分類に対する各公報の二値分類を実施した結果、図11に示す通り（a）再現率と（b）適合率ともに、各ツールで70%程度（大分類2では90%超<sup>29)</sup>）の精度が得られた。

##### 2) 二値/多値分類, 単一/複合教師による差

同様に、先述したツール $\delta$ で単一教師による二値分類を行った結果に対し、単一教師による

多値分類を行った結果、および複合教師による多値分類を行った結果をそれぞれ比較したものを図12に示す。

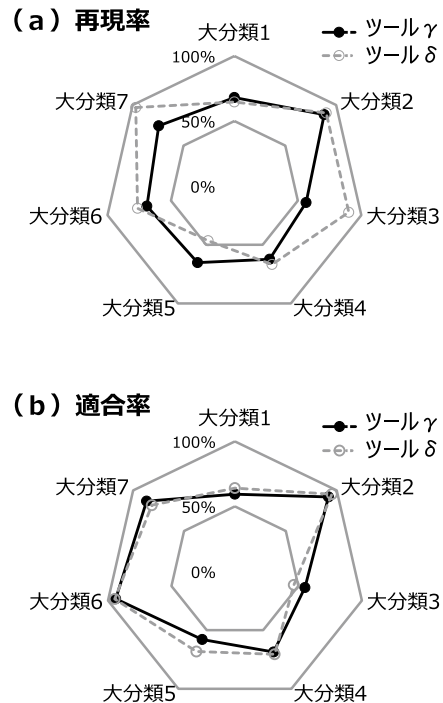


図11 ツール別での分類付与結果

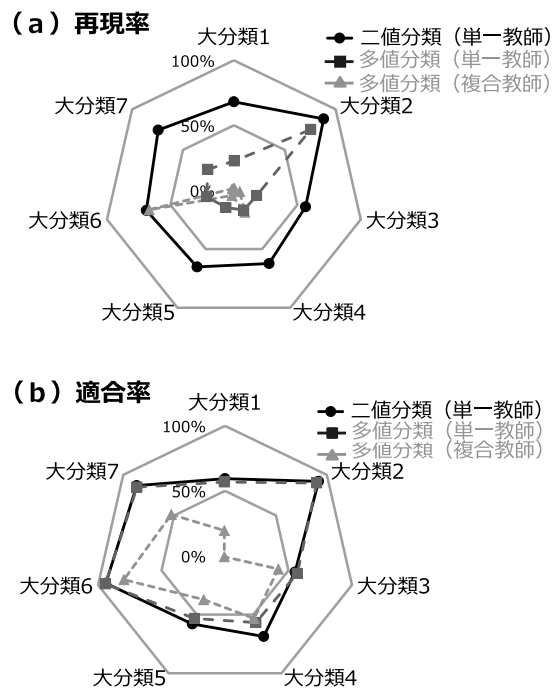


図12 二値/多値分類, 単一/複合教師別での分類付与結果

この結果から (a) 再現率と (b) 適合率ともに、多値分類よりも二値分類の方が、また複合教師よりも単一教師の方が、有利な結果となることが確認された(複合教師における結果は、いわゆる「醜いアヒルの子定理<sup>30)</sup>」によるものと推測される)。

紙幅の都合により本稿では詳細な考察は割愛するが、このような分類付与を行う上での教師データの最適化に関しては、2019年度情報検索委員会第2小委員会第2ワーキンググループが、より多岐にわたる検討を行っており、別途本誌にて報告を予定しているため、それも参照されたい。

## 5. おわりに

以上、ここまで述べてきた無効調査および動向調査における各種ツールの実力検証と活用検討の結果を表10に纏めた。今回の検討により、各業務における効率化や精度向上に向けた各種ツールの選定および活用手法の考え方について、ある程度参考になる情報が得られたのではないかと考える。

また、本研究は特許調査の効率化を主な目的

として行ったものであるが、上記提案した手法により各種ツールの機能を活用することで、検索式の構築や抽出特許の精査といった人手による作業を省略することができれば、業務効率化のみならず調査担当者の検索スキルの差による調査レベルのバラつきの緩和にも繋がると考えられる。したがって業務品質の安定化や、あるいは特許調査に不慣れな部門(開発や営業等)の担当者による特許調査を支援する面でも有用であると期待される。

なお、念のため申し添えるが、本研究は各種ツール間での優劣を比較することを目的としたものではなく、ツールの仕様の違いによりどのような条件下ではどのような傾向が出るのか、反対にツールによらず現れる傾向があるのか、調査の目的を達成する上でどのような点に留意すべきか等の検討から、目的に応じたツールの選定やその使い方の提言を目指したものである。

そして得られた結果もあくまで今回のデータを今回の条件で解析した結果であるため、データや条件の変更、あるいはツールの進化等により傾向は変わりうる。したがって当然ながら特定のツールを推奨または批判するものではない。

表10 本研究結果のまとめ

	無効調査	動向調査
ツールの 実力検証	<ul style="list-style-type: none"> <li>・基本的に再現率と適合率のトレードオフ(抽出件数上限の無い特許検索ツールでの再現率は約80%)</li> <li>・機械分野で有利な傾向</li> <li>・類似度順500位以内ではほとんどの特許検索ツールにおいて再現率は約40%で同等</li> <li>・レーティングツール併用にて適合率向上(教師データを多数必要とするレーティングツールは不利な傾向)</li> </ul>	<ul style="list-style-type: none"> <li>・母集団約3,000件に対し教師データ100件での二値分類においては、人手による調査結果に対し70%程度の正解率(分野次第では90%超)</li> <li>・精度は二値分類、工数は多値分類がそれぞれ有利</li> <li>・分類付与における教師データ(正解教師)には、付与させたい分類以外の特徴量をなるべく含まない単一教師を用いることが有利</li> </ul>
ツールの 活用検討	<ul style="list-style-type: none"> <li>・高再現率の特許検索ツールでの類似検索結果を母集団に、少数の教師データによる解析が可能なレーティングツールを用いた並べ替えを推奨</li> <li>・教師データを変更した複数パターンの並べ替え結果を横断的に確認することで効率向上の可能性あり</li> </ul>	<ul style="list-style-type: none"> <li>・高精度が期待できる分野では有用</li> <li>・二値分類/多値分類の選択可否、単一教師の準備可否など、ツールや状況により適否検討が必要</li> <li>・高精度化には教師データの最適化検討が必要(当小委員会第2ワーキンググループの報告参照)</li> </ul>



本稿が会員企業各社にて特許調査の効率化を検討する上で多少なりとも参考になれば幸いである。

## 注 記

- 1) 麻川倫広, 「特許庁における人工知能技術の活用に関する取組について」, Japio YEAR BOOK 2018, p.7 (2018) 表1より一部抜粋(省略を意味する二重破線は執筆者が追記)。  
[https://www.japio.or.jp/00yearbook/files/2018book/18\\_a\\_01.pdf](https://www.japio.or.jp/00yearbook/files/2018book/18_a_01.pdf)
- 2) ベンチマークとしての客観的なデータとして、無効調査では実際の無効審判事例があり、動向調査では日本特許庁による「特許出願技術動向調査」等が挙げられる。また、導入へのハードルについては、無効調査は目的とする先行文献すなわち対象特許の新規性および/または進歩性を否定できる文献を見つけさえすれば良いという点で、実施前のクリアランス調査のような漏れのない検索が求められる訳ではなく、動向調査に関しても、マクロ分析のような概略の傾向が把握できれば良い場合には、同様に多少の漏れは許容されうるという点で、「比較的」ハードルが低いと表現した。
- 3) 野崎篤志, 「特許情報をめぐる最新のトレンド」, Japio YEAR BOOK 2018, p.61 (2018) 表2より引用。なお、2020年3月現在、ツール名「IP Samurai」とベンダー名「ゴールドアイピー」はいずれも「AI Samurai」に改称されている。  
[https://www.japio.or.jp/00yearbook/files/2018book/18\\_a\\_08.pdf](https://www.japio.or.jp/00yearbook/files/2018book/18_a_08.pdf)
- 4) 間瀬久雄, 「特許を対象とした概念検索の技術課題」, Japio YEAR BOOK 2010, p.202 (2010)  
[https://www.japio.or.jp/00yearbook/files/2010book/10\\_3\\_08.pdf](https://www.japio.or.jp/00yearbook/files/2010book/10_3_08.pdf)
- 5) 例えば「案件A, Bの2件について、それぞれ母集団として100件ずつの検索結果を得て、当該母集団の確認において、案件Aでは目的の特許を発見できたが案件Bでは目的の特許を発見できなかった」という場合、一般的な「適合率」は全検索結果の200件を分母に用いて求められる0.5% (1件/200件) となるが、本研究の「補正適合率」は目的の特許を発見できなかった案件B (発見不可案件) を除外し、案件Aのみを考慮して1% (1件/100件) として算出する。すなわち発見できたか否かについては「再現率」で表し、「補正適合率」は発見できていることを前提として発見にかかる工数を評価するという変更を行っている(「補正適合率」については注記6) および注記12) も要参照)。
- 6) 通常、適合率は図1の通り「全検索結果(検索結果の母集団の全特許件数)」を分母として算出するものだが、本研究では、後述する「レイティングツールによる並べ替え」において、並べ替えの前後で調査に要する工数がどのように変化するかを比較するため、発見したい特許に到達するまでに「実際に確認した特許の件数(実質調査件数)」を分母として設定している。例えば、検索結果の母集団100件中に発見したい特許が1件含まれているというケースで、当該母集団の確認において先頭から10件目で目的の特許に到達した場合、図1によれば「適合率」は母集団の100件を分母に用いた1% (1件/100件) となるが、「補正適合率」は実質調査件数の10件を分母に用いた10% (1件/10件) として算出するという変更を行っている(「補正適合率」については注記5) および注記12) も要参照)。
- 7) 日本知的財産協会2015年度特許第1委員会第4小委員会, 「無効審判事件分析による特許審査の質の検証」知財管理, Vol.67, No.1, pp.40~50 (2017) にて検討対象としていた171件の事例から証拠に採用された先行文献が非特許文献や外国特許文献のみのもの等を除外し選定。上記の理由は、本研究で用いたツールには非特許文献や外国特許文献の検索ができないものも含むことから、可能な限り条件を揃えてツール間の比較を行えるようにするためである。なお、この点から本研究の評価に用いる母集団の検索においては、外国特許文献の検索が可能なツールに関しても、日本に出願された特許のみに絞って行った。
- 8) 「新規性」に分類した案件は、無効理由として特許法第29条第1項第3号または第29条の2の少なくともいずれかを有するもの(第29条第2項を合わせ持つものも含む)を意味し、「進歩性」に分類した案件は、無効理由として第29条第2項を有し、かつ第29条第1項各号および第29条の2のいずれも有さないものを意味する。なお、「新規性」、「進歩性」のいずれの案件も、第36条など上記以

- 外の無効理由を合わせ持つものも含む。
- 9) 表5には、利用目的によって「スクリーニングツール」、「自動分類ツール」、「ノイズカットツール」等とも称されるものを含むが、本研究における利用方法から、ここでは「レーティングツール」と称する。
  - 10) ツールによって「概念検索」、「コンセプト検索」等とも称されるが、キーワード（単語）の組合せのみでなく文脈上の意味を考慮して検索を行う機能を指す。
  - 11) 検索対象範囲は注記7)に記載の通り、日本特許に限定し、また対象特許の出願日以前（対象特許が優先権主張出願の場合は優先日以前、分割出願の場合は親出願の出願日以前）のものに限定した。なお、厳密には特許法第29条の2を考慮し出願日前1年6月以内の出願人同一または発明者同一特許を除外すべきではあるが、本研究ではこれは考慮していない。
  - 12) 本研究の「補正適合率」においては、通常の「適合率」に対し、注記5)で説明した発見不可案件の除外に加え、注記6)で説明した実質調査件数での分母設定という変更を行っているため、対象特許50件のうち目的とする先行文献を発見できた案件のみを対象とし、さらに検索結果の母集団件数そのものを分母とするのではなく、当該母集団の何件目で目的とする先行文献を発見できたか、すなわち類似検索機能または並べ替え機能の結果における「類似度の順位」を実質調査件数として分母に設定し算出した。例えば進歩性を無効理由とする対象特許A（発見したい先行文献はa1とa2）、B（同、b1とb2）、C（同、c1とc2）について、それぞれの類似検索の結果として100件ずつの母集団における類似度順で、a1が10件目、a2が20件目、b1が10件目で発見され、b2、c1、c2はいずれも発見されなかった場合、「適合率」としては全体で300件の母集団中にa1、a2、b1の3件の先行文献が含まれていたことから1%（3件/300件）と算出されるが、「補正適合率」としては、対象特許Aに関しては母集団100件のうち上位20件の特許を確認することでa1とa2の2件を発見でき、対象特許Bに関しては母集団100件のうち上位10件を確認することでb1の1件を発見でき、発見不可案件のb2、c1、c2は除外するため、実質調査件数30件中に3件の先行文献が含まれることから10%（3件/30件）と算出される。なお、現実の調査では「目的とする先行文献（正解）」があらかじめ分かっている訳ではないため、「b2、c1、c2が母集団中に存在しないこと」は母集団100件ずつの全てに目を通さなければ分からないことであり、対象特許Aについても「上位20件の調査でa1、a2の2件が見つかったとしても、より良い先行文献がないか念のため母集団100件を全て確認するはず」と考えれば、分母を300件として「適合率1%」で評価する方が適切とは思われるが、本研究の無効調査において通常の「適合率」で評価すると、特に抽出件数上限の大きいツールでは、1件でも発見不可案件がある場合、分母に数万件といった極端に大きい値を設定することで適合率がごく小さい値となり、ツール間の差がほとんど見えなくなってしまうため、ツールごとの「類似度判定の的確さ（ひいては目的とする先行文献の発見にかかる工数）」をより見えやすい形で比較する意味で「補正適合率」を用いることとした。
  - 13) 日本技術貿易株式会社IP総研、「自動無効資料調査機能検証」（2017）  
関連記事：中岡浩，「【生かせ！知財ビジネス】NGB、インド発のAI特許ツールを検証」, SankeiBiz (2017)  
<https://www.sankeibiz.jp/compliance/news/171102/cpd1711022226010-n1.htm>
  - 14) PLASDOC オンライン研究会，「人工知能（AI）を用いた化学分野における無効資料調査・先行技術文献調査への活用検討」, 情報の科学と技術, Vol.68, No.9, pp.470~476 (2018)
  - 15) 類似検索機能において無効審判の情報をも参照していた場合、その先行文献は検出できて当然となるため、ツールの実力を評価できないことが懸念された。この影響の有無は完全には確認できなかったが、審査時の引用文献は参照されても、無効審判時の先行文献は参照されていないと推測されたこと、及び本研究の趣旨はツールの優劣を比較することではなく、各種ツールの活用方法の検討にあることから、この点の追究は行っていない。
  - 16) 確実に無効化しなければならない案件では、母集団が何万件であろうと確認すべきケースもあると思われるが、経験的に「これだけ調査して見つからなければ先行文献は発見困難」と推測

される場合や、費用対効果面でそれ以上工数が割けない場合など、調査担当者ならばある程度の見込をもって調査を行うものと推測される。

- 17) 無効調査の母集団として何件程度が妥当であるかは、企業ごとや案件ごと、あるいは調査担当者の人数等により異なると思われるが、本研究を実施した委員の経験上、平均的に「数百件から数千件」との感覚であったことから、目安として500件を設定した。
- 18) レイティングツール欄および特許検索ツール欄の記号は、表4および表5の各種ツールに対応している。
- 19) ツールの仕様により、一部条件について横並びの比較ができていない箇所はあるが、なるべく近い条件となるよう設定した。
- 20) 表6に示す通り、教師データや解析フィールドを完全に揃えての比較にはなっていないことが影響している可能性もあるが、ツールの仕様上これを揃えることはできないためここでは考慮しない。
- 21) 教師データは全請求項、解析フィールドは類似度Top10段落に揃えて比較した。
- 22) 表6に示す通り、解析フィールドも類似度Top10に揃えて比較した。
- 23) ツール  $a$  では各請求項や各段落に対し個々に類似度スコア付与を行うことが可能なためこの機能を利用したが、ツールによってはこのような指定はできなかつたり、人手により請求項や要約等を抽出したデータを用意する必要があつたりと仕様異なる。教師データや解析フィールドに用いる項目の範囲については、別途2019年度情報検索委員会第2小委員会第2ワーキンググループが検討しており、今後本誌にて報告予定であるため、そちらを参照されたい。なお表6に示す通り、ここでは教師データを全請求項に揃えて比較した。
- 24) 複数の教師データとして、例えば「請求項」と「効果」などのように、発明の特徴を含みつつ異なる観点での記載が想定される範囲をそれぞれ用いると、並べ替え後の各集合内の類似度順に差が生じ

やすくなり、そのようなアプローチの異なる集合を横断的に確認することで、いずれかの集合で目的とする先行文献をより早く発見できる可能性が高くなるものと推定される。また、ある集合で既に確認した特許を他の集合で再度確認することによるロスを避けるためには、RPA等を活用して、ある集合から抽出した特許が他の集合の確認済の類似度順範囲に含まれている場合はそれを除外する、といった操作を行うことも推奨される。

- 25) 特許庁、「特許出願技術動向調査 平成28年度ファインバブル技術」  
[https://www.jpo.go.jp/resources/report/gidouhoukoku/tokkyo/document/index/28\\_10.pdf](https://www.jpo.go.jp/resources/report/gidouhoukoku/tokkyo/document/index/28_10.pdf)
- 26) 動向調査において、教師データおよび解析フィールドは、いずれも「名称+要約+全請求項」を設定した。
- 27) すなわち、この動向調査における「適合率」は図1の定義に沿って算出されたものであるため、本章では「補正適合率」の語は用いない。
- 28) 正解教師の50件全てを単一教師とすることはできなかった大分類もあるが、いずれの大分類でも正解教師の80%以上が当該大分類のみを有する単一教師となるように選択した。複合教師の考え方も同様である。
- 29) 表9の通り、大分類2は「ファインバブル計測技術」となっており、内容として機械分野に近い技術であることから、無効調査での分野による傾向と同様に有利な結果が得られたものと推測される。
- 30) 詳細は省略するが、「醜いアヒルの子を含む  $n$  匹のアヒルがいるとすると、醜いアヒルの子と普通のアヒルの子の類似性は、任意の二匹の普通のアヒルの子の間の類似性と同じになる（多数の観点での特徴量を持つ場合、本来判別したい観点以外の特徴量における類似性の影響を受けてしまうことにより、適切な判別ができなくなる）」という理論。

(URL参照日は全て2020年3月12日)

(原稿受領日 2020年7月20日)