

概念検索を利用した 無効資料調査

六車技術士事務所長 **六車 正道**

PROFILE

約 36 年間、日立製作所において特許情報の活用企画と実務を担当し、2006 年 2 月に技術士事務所を開設し、特許情報の活用促進に関するコンサルタントに従事。講演や著作多数。PatentCity の運営者。技術士（情報システム）

✉ <http://www.patentcity.jp/mailtomug.htm>

TEL 050-8012-2416

〔要約〕概念検索は絞り込んだ技術の検索に適しており、無効資料調査にも適している。やり方は、概念検索によりヒットを含む可能性の高い集合を作り目視チェックする、と共に信頼性保証のために数千件の特許も目視する。この利用法は、概念検索を予備検索としての位置付けではなく「コア検索」として位置付けるべきであろう。

1 はじめに

無効資料調査では調査漏れのできるだけ少ない高精度の調査が求められる。一方、特許出願前などの簡単な先行技術調査では、かけられる費用が少ないので、手抜きをするわけではないが一定の漏れを許した調査になっている。つまりごく大雑把に言えば、高精度の無効資料調査とは多くの費用をかける調査ということができない。その結果、無効資料調査では請求範囲のフレーズごとに比較検討するような詳細な調査まで行なわれることが多い。

ところで、いかなる調査であってもできるだけ効率的な手法をとることが求められる。費用が比較的潤沢である無効資料調査であっても、効率の悪い方式で長時間を費やして、努力したという証拠にするようなものであってはならない。

最も高精度であるが最も効率の悪い調査とは、全資料（1年間の日本公開特許だけで 25～30 万件）を順次目視チェックするものである。膨大な費用をかけて、眠気の誘惑に打ち勝ってこの調査をもし完遂できれば絶対に漏れない調査にできるはずである。

効率の良い調査とは、例えば、数件のヒット（当たり）を含むと思われる 100 件程度の集合まで検索式で絞り込んでそれを目視チェックする方法である。しかし、あまりに数少なく検索で絞り込むと漏れの生じる恐れがある。そこで多くの無効資料調査では、検索では数千件程度の粗い絞り込みにとどめておき、それを目視チェックすることになる。一般的には目視チェック件数が多いほど費用がかかるが漏れの少ない調査になる、はずである。

検索の方法として、ヒットを含む可能性の高いIPC（国際特許分類）やFI記号、Fターム、またはキーワード、場合によっては出願人名などで絞り込むのは周知のことである。ところが、これらをうまく使いこなすのは容易なことではない。上手に絞り込んだ検索のつもりが漏れを起こしてしまい、労力を使っただけで目視チェックを台無しにすることも時折発生する。

最近、概念検索という検索手法が利用されつつある。概念検索は質問文の中のキーワードにシステムが自動的に重み付けを行ない、それに沿ってデータベース中の特許に類似度の点数を与え、点数の高い方からリストアップするものである。これをうまく使うことができれば、類似度の高い特許の集合を簡単に作ることができ効率的な調査に貢献する。ところが、上手に使いこなす方法が広く知られている状況になく、概念検索を無効資料調査に使うのは無謀であるという考えの人もある。

本稿では、無効資料調査に概念検索をどうすれば上手に使えるかについて考察する。

2 無効資料調査の特徴

無効資料調査（無効化資料調査ともいう）はある特許の最後のステップの調査である。つまり、開発着手前の動向調査や特許出願前の先行技術調査、製品出荷前などの侵害予防調査であれば、もし調査漏れがあったとしても次のステップで対策をすれば良い。しかし無効資料調査はその特許の有効性を争う最後の段階であり、次のステップは無い。その特許が有効かどうかにより高額な賠償金を伴うものであり、出願前の簡単な調査に比べて数倍から数十倍の費用をかけることも多い。

なお、対象の特許を無効にできそうな有力な公知例が見つかったならばそれ以上の調査を続ける必要はなくなる。また、調査の目的とする技術は請求範囲に記載されていることであり明白であり、多くの場合狭い範囲である。一方、特許調査の対象範囲は広がる特徴がある。なぜならば、すでに特許審査とかいろんな機会に調査されているので、むしろズバリの該当技術の範囲でない部分の調査が期待される面があるからである。

無効資料調査の特徴を明確にするために侵害予防調査の特徴を説明する。侵害予防調査では、1つの類似特許が見つかってそれ以外にもあるかも知れないので、基本的には特許の権利期間の全体を調査する必要がある。ただし、自社特許は調査不要である。また、自社製品のどの部分が他社特許を侵害するか分からないので、調査の目的、つまり特許調査すべき技術範囲を決めること自体が難しい。上手に絞らないと広範囲の技術について調査が必要になってしまう。逆に調査対象はズバリの該当技術の範囲にすることが多い。

3 データベースと検索式による調査

3.1 データベース検索を利用した調査

1990年代前半までは、無効資料調査などにおいては紙資料による全文明細書の手めくり調査が必須とされ、データベース検索は不適切とされていた。しかしそ

の背景にはデータベース利用料金が検索回数や出力量にリンクしており高額であったこと、また参照できる資料も限定的であったことがある。つまり、検索に続いて直接見られるのは抄録や代表図面だけであり、それらで粗ぶるいして絞り込み、その後別の手段で明細書を見るところがあった。その後データベースの出力として明細書を見られるようになったがしばらくの間は1件を表示させるのに数十秒など時間がかかる間は似たような位置付けであった。

ところが、データベース検索の改良が進み、利用料金が大幅に下落し、明細書の表示が2、3秒でできるようになったことで、データベース検索でも全文明細書による調査が可能になった。

現在においては、IPCやF1記号による検索で数千件程度に絞り込んで抄録や明細書意を見て調査するやり方は、かつての紙の明細書を使った調査と全く同一の機能を果たしているといえる。

ところがデータベース検索は便利なものであり、条件によってはいくらかでも件数を少なく絞り込める。目視チェックする件数は少ない方が時間は短くて済むので、費用を安くしたい場合にはキーワード検索により100件くらいまで絞り込むことも行なわれる。ところが検索テクニックが稚拙であったり、絞り込み過ぎることで漏れが多くなり、「データベース検索は無効資料調査には適していない」と誤解を生むこともある。

3.2 優先順位を設けた効果的な調査

100件や200件まで絞り込むと短時間で明細書などの目視チェックを終えられるが、絞り過ぎのために漏れがありうる。そこで、例えば3,000件も目視チェックするほどの費用はかけられないという場合、次のようなやり方がある。

まず、漏れの少ない可能性の高い集合を検索式で作成し、例えば3,000件あったとする。次にこれを母集合として、最も可能性の高いキーワードなどでさらに絞り込み検索を行なってヒットの可能性の高い100件前後に限定し、それを先に目視チェックする。次に残りの特許を対象に、2番目に可能性のあるキーワードなどで数



百件に絞り込んで目視チェックする。残りの特許も最終的には目視チェックする。このように優先順位の高い集合とそうでないものに分けることで調査の綿密さを調整でき、効率的でありながら高精度の調査を実現できる。

調査の綿密さとは、例えば明細書をどの程度綿密に見るかということである。つまり、ある特許明細書の先頭部分を見て全く異なった内容である場合、それ以降の部分を見るか見ないかということである。このとき、一見ただけでは全く異なると思える場合に何か違和感を感じて明細書の細部まで見ていると思わぬ公知情報を発見することを経験することがある。細部を見るか見ないかの判断は慎重に行なう必要がある。

なお優先順で分けるのは3段階に限ったわけではなくそれ以上に分けることも実際的である。

ところがこのように目視チェックする程度に差をつけると、3,000件を目視チェックしたと言えないのではないかと異論があるかもしれない。確かに3,000件の全件を「明細書の細部まで見た」とは言えないが「明細書も一応見た」とは言えるであろう。もちろん、3,000件の細部まで目視チェックするほどの膨大な費用をかけられる場合には、当然のことであるがそうすべきである。

ところで、優先順位の高い特許の集合を簡単に作る方法として概念検索がある。

4 概念検索とは？

概念検索とは次のようなものである。

- ① 検索目的とする技術の説明文を質問文として入力する。
- ③ 回答としてヒットの可能性の高い順に特許が表示される。

4.1 概念検索の特徴

概念検索を上手に使うための特徴については既に多くの記事¹があるが、代表的なものとして次のようなものがあげられる。

- ① IPC や F I 記号、またキーワードなどで絞り込むための検索式を作る必要がない（検索式と組合せた概念

検索も可能）。

- ② 質問文中のキーワードに自動的に重み付けが行なわれて処理される。
- ③ その重み付けは質問文での出現回数と、そのキーワードを含んでいるデータベース全体での特許件数から計算される。現在の概念検索では文法は利用されない。
- ④ その重み付けにそってデータベース中の各特許の得点を計算し、点数の高い特許を類似度の高い特許として表示する。
- ⑤ 絞り込んだ内容の検索に適している。

4.2 絞り込んだ検索

概念検索は質問文で表わされる技術内容に絞り込んだ検索が可能である。この質問文がキーワード2、3個のあまりに短い質問文では絞り込みが浅く、該当する特許は膨大な数になる。例えば質問文が「放射性物質の回収」に適合する特許は膨大な件数になるが、「土壌中のセシウムなどの放射性物質を多孔性物質による吸着で回収する」は数少ないものに絞り込まれる。もし後の内容の特許を概念検索したいのであればこの文章を質問文として使える。

一方、長い文章でも関係技術の説明まで広げた文章とか、質問したい技術と関係の薄いキーワードを何度も繰り返すような文章では絞り込んだ質問にならない。特許明細書は関係技術まで説明し過ぎであり、請求範囲は多くの場合、質問技術と関係の薄いワードの繰り返しが多い。「要約」も関係技術まで広げた文章となっている場合が多い。

通常、30～80文字程度の文章であれば適切な長さの質問文になることが多い。なおこの範囲から少々外れていても問題ない。また、長いキーワードを多く含む場合はもう少し長い文章が良い。

4.3 上手な概念検索のやり方

これまでも書かれている¹ので、ここでは代表的なものの紹介にとどめる。

- ・ 質問文は技術範囲を明確に特定した短文が良い。
- ・ 質問文は具体的なキーワードを使った文章が良い。

- ・ 質問文は技術的表現だけでなく、目的や効果を含めることも有益。
- ・ 質問文には同義語はあまり要らない。場合によっては邪魔になる。
- ・ IPC、FI 記号、F ターム、キーワードや発行年などによる絞り込みを併用した検索は有益。
- ・ 特許番号指定の概念検索は簡単で便利だが漏れは多い。
- ・ 絞り込まれた質問文の場合、回答は上位 50 件（～100 件）まで目視チェックするのが良い。
- ・ 質問文を変えて概念検索を数回行う。これは必須であり常套手段といえる。

5 概念検索による無効資料調査

ここでは実例を使って無効資料調査への概念検索の利用の検討を行なう。

5.1 無効資料調査の実例

次のような内容の特許を無効にできる特許を探す調査を考えてみる。

テーマ；複数機関への手続き代行システム

説明 1；ネットワーク経由で複数の機関に電子的な手続きをおこなう代行システムで、フォーマットを特徴づけるデータの類似度を計算してデータベースから検索する。

説明 2；地方自治体や電力会社など複数の提出先にネットワーク経由で電子的な手続きの代行システムで、異なる機関の手続きに必要なフォーマットをデータベースから検索して記入。

特許請求範囲；クライアント端末画面から入力された第 1 の機関への申請届け出情報を第 2 の機関への申請届け出に利用するシステムにおいて、複数機関への申請届け出の各フォーマットの各項目を特徴づけるデータを保有する手段と、その各項目を特徴づけるデータの類似度を計算して第 2 の機関への申請届け出に必要なフォーマットを検索する手段とし、その情報に基づいて第 2 の

機関への申請届け出フォーマットへの記入をガイドする手段を有する申請届け出代行システム。

概念検索の実施；まず「説明 1」の文章で概念検索を行ない 35 件目にやや近い内容の特許が見つかったが細部は異なるものであった。この結果を受けて細部を変えていくつか検索し、その後「説明 2」の文章で検索し、近いもの数件を得た。また特許請求範囲の文書でも概念検索を行なって比較検討した。

表 5-1 は「説明 2」の文章を質問文として概念検索を行なった結果である。発明の名称から見ると似たものが見つかっていることが伺える。

表 5-1 「説明 2」文による概念検索 システム：Shareresearch

項番	スコア	公報番号	発明の名称	出願人
1	158	特開2003-085263	手続き代行方法およびシステム	株式会社日立製作所
2	144	特開2002-236739	個人の住所変更を関連機関に通知...	株式会社ジネット
3	135	特開2001-357043	法令様式デジタル大全集	株式会社日本法令
4	135	特開平11-328288	電子申請・電子帳票流通サービ...	大日本印刷株式会社
5	132	特開2002-063393	多機関手続き仲介方法	塩崎久美
6	131	特開2007-172425	住所変更届及び返信用封筒の提供...	李載溶

表 5-2 は、この検索結果の上位 50 件を目視チェックして、本案件の特許請求範囲のフレーズごとに一致する記載があるかどうかを検討した結果、その他である。4 件目の特開 11-328288（以下 288 特許という）はかなり近い内容である。しかし 288 特許では「申請手続きの帳票の属性を保存して検索に利用する」とし「帳票ごとに重み」を付けて管理しているが、本案件の請求範囲のように「類似度を計算」とまでは書いてないので△とした。

なお、「説明 2」の質問文で使っている次のキーワードは 288 特許には一度も書かれていない。しかし 288 特許が上位に出てきていることは概念検索の最も興味ある点である。

- ・ 地方自治体
- ・ 電力会社
- ・ 代行
- ・ フォーマット

5.2 特許請求範囲で検索できるか？

表 5-3 は本案件の特許請求範囲「クライアント端末画面から・・・」で概念検索を行なった結果である。

表5-2 特許請求範囲のフレーズごとの対応表 ※ 実物とは一部相違あり

特許請求範囲	質問文	説明2		説明1	請求範囲
	順位	4 件目	1 4	3 5	4
	特許番号	特開平 11-328288	特開 2006-221493	特開 2001-209693	特開 2006-221493
クライアント端末画面から入力された第1の機関への申請届け出情報を第2の機関への申請届け出に利用するシステムにおいて、		○	○	○	○
複数機関への申請届け出の各フォーマットの各項目を特徴づけるデータを保有し、		○	○	○	×
その各項目を特徴づけるデータの類似度を計算して		△	×	×	×
第2の機関への申請届け出に必要なフォーマットを検索し、		○	○	×	×
その情報に基づいて第2の機関への申請届け出フォーマットへの記入をガイドする申請届け出代行システム。		○	○	○	○

288 特許は 105 件目に出ているが、もし上位 50 件(また 100 件目) までしか目視チェックしていなければ見つけることは出来なかった。なお、筆者の経験によればこの結果は「概念検索では上位 150 件まで目視チェックする必要がある」と理解すべきではなく、「請求範囲を質問文に使うのは(多くの場合) 不適切」とすべきと考える。

288 特許が上位に出てこない理由は、簡単に言えばこの特許の明細書が質問文と類似の書き方になっていないということである。具体的に比較すれば、228 特許の明細書や請求範囲は次のような書き方をされている。

228 特許の書かれ方；・・・ユーザ端末に電子的ネットワークを介して接続されたユーザ支援ナビゲーション部と、ナビゲーション部にデータの送受信が可能に接続された帳票データベースとを備えた電子申請・電子帳票流通サービスシステムであって、帳票データベースには、各種申請手続きに使用する帳票と、これら各手続きに関連する帳票と、それらを互いに関連付ける属性に関するデータが格納され・・・

これに対して質問文(本案件の特許請求範囲)は先に示した通り「クライアント端末画面から入力された第1の機関への申請届け出情報を第2の機関への申請届け出に・・・」となっており、228 特許とかなり異なる文章であるため類似度が低くなったものと言える。

表5-3 特許請求範囲による概念検索

項番	スコア	公報番号	発明の名称	出願人
1	302	特開2005-275904	申請手続き支援システム	セイコーエプソン株式...
2	268	特開2002-063393	多機関手続き仲介方法	塩崎久美
3	239	特開2003-150691	自動車登録申請に係る代行依頼シ...	有限会社ミナミ情報シ...
4	239	特開2006-221493	電子申請方法	キヤノン株式会社
5	235	特開2002-092142	住所変更方法及び住所変更システム	日本電気株式会社
101	194	特開2002-123592	申請請負システムおよびサーバシ...	旭化成ケミカル株式...
102	193	特開2011-039799	電子記録復権申請システム	株式会社野村総合研究所
103	193	特開2003-216784	電子申請システム及び電子申請方法	株式会社社会情報クリ...
104	193	特開2002-083255	電子フォーム利用システム	大日本印刷株式会社
105	193	特開平 11-328288	電子申請・電子帳票流通サービス...	大日本印刷株式会社

一方、表5-3の上位に出てくる多くの特許は本案件と異なるのになぜ上位に出てくるのだろうか。1 件目の特許を見てみると次のような記載が繰り返し出てくる。太字は質問文の請求範囲に含まれるワードである。

「第2発明は、第1発明において、上記申請用端末装置には上記**手続機関**が**複数**提示されており、**複数の手続機関**から**申請を必要**とする**手続機関**を選択できることを・・・」

このように一致するワードが多いので、内容的には異なる特許であるが上位に出てくるものである。2 件目、3 件目も概ね同様である。4 件目の特開 2006-221493 は「申請情報と帳票フォーマットを紐付ける(関連付ける)」となっておりやや近い感じがするが「各項目を特徴づけるデータを保有する」とは書いてない。(⇒表5-2)

5.3 高精度の調査の保証

概念検索で近い内容の特許を探すことに成功しても、それで終わりにできないことが多いと思われる。なぜな

ら、もっと調査に時間をかければもっと近い内容の特許が出てくるかもしれないからである。

一般的なことであるが、予算が50万円であった場合、4,000件くらいの特許を目視チェックすることが期待される。そのような場合、概念検索で短時間に数件の近い特許を見つけてもそれで終わりにすることはできない。高精度の特許調査を保証するのであればやはり4000件の目視チェックを行ない類似の特許を探す必要がある。

その時、概念検索で50件や100件を目視チェックした集合がいくつかあればそれを合計して例えば400件の集合を作り、4000件の集合からそれを除いて目視チェックするのが効率的なやり方である。ただしこの場合、 $4000-400=3600$ 件とはならない。同じ特許があった場合のみ除かれるので多くの場合3800件程度にしか減らない。また、このやり方は概念検索の集合を検索式で利用できるシステムでのみ利用可能である。

その4000件の集合を作る検索式では、概念検索で見つけた近い内容の特許から得た知識を生かすことは当然のことである。また、検索式の4000件から概念検索の400件を除いても3800件にしかならないということは、概念検索が検索式から外れたところを数多く見ていることを意味する。つまり概念検索は検索式にとられない広い領域を検索していることを意味し、これも概念検索のメリットである。

なお、このことは実は特許調査の根幹にかかわる重要な問題を含んでいる。つまりIPCやF I記号、またキーワードによる検索では、特に無効資料調査などの場合は、上手に絞り込むことが非常に難しいということである。(その結果、3000件とか4000件もの膨大な特許を目視チェックすることになる。)

検索式で絞り込むことが難しい例として288特許を見ても、この特許に付いているIPCはG06F19/00、3/00、13/00、17/30である。17/30は「情報検索:そのためのデータベース構造」とやや限定的であるが、それ以外は/00であり大雑把なIPCである。IPCで絞るのは簡単ではないことが伺える。またF I記号は5つのうち4つまでが途中で変更になっており現在は使われ

ていない。継続しているG06F13/00,354Dは「端末における入出力制御」というものでそれほど絞り込みに役立つものではない。

なお、その3800件を目視チェックする場合も、3.2で紹介したように優先順位を設けて重要なものから目視チェックするのが効率的で高精度の調査法といえる。

以上の検討は費用が十分かけられる場合のことであり、もし費用を1/2程度にしたいと言う場合には検索式で目視チェックする件数は2000件程度で終わることになり、概念検索の役割はいっそう重要なものになる。

簡単な概念検索だけで無効資料調査の精度を保証することはできないが、上手な質問文で何度か概念検索をやり直し、優先度の高い集合を集中して先に見るといった調査のやり方は、効率的で高精度の特許調査に貢献する。このような概念検索は、検索式作成のための予備検索ではなく、重要な部分の検索という意味で「コア検索」と言うべきであろう。

6 おわりに

無効資料調査において概念検索は役立つ結果を期待できる。それは概念検索が焦点の絞れた検索に適しているからであり、検索のやり直しが比較的簡単であるためと言えよう。

概念検索は簡単な動向調査には役立つだろうが無効資料調査には使えないと理解している人は、実は全く逆であると認識を改めるべきである。

なお本稿では説明しなかったが、概念検索を行なう人は目的技術の知識が豊富であった方が良い結果を得ることが期待できる。さらに、絞り込み検索などのテクニックを使うことでいっそうの再現率の高い検索、高精度の特許調査が期待できよう。

参考文献

- [1] 概念検索による特許情報の活用法、六車正道、発明、Vol.100、No.4、2004年4月