

概念検索の使いこなしに関する論文

六車 正道*

概念検索の使いこなしに関する論文はすでに 30 編近く発行されている一方で、概念検索は再現率が低く実務には使えないという意見もある。本稿でこれらの各論文の概略を紹介した。これらを見ると、検索実験には実験のテーマ数が数十件など大量に実施されたものと少ないものがある。テーマ数が多い場合、質問文は請求範囲などが使われている。テーマ数が少ない場合、質問文は意志を反映させて利用者が作った短文を使い、ヒット判定に使う目標特許群は綿密な調査をして設定しているものが多い。これらにより、なぜ概念検索の利用で再現率が高いものと低いものがあるのか推察される。

キーワード：概念検索，特許調査，情報検索，データベース，キーワード分布，出現頻度

1. はじめに

情報は「受け手の知識を変更する」ために必要なものと言われる。一方、特許情報は新技術を平均的な技術者が実施できるように詳細に書かれた技術的文書であり、全技術を 1 つの体系のもとに発行される便利なものである。また出願された特許は、(一部の取り下げられたものを除いて)すべてが 1.5 年後には公開公報として発行され、データベースで参照可能になっている。

その特許情報も利用者の知識を変更するために使われる。例えば、自分の発明が特許になるか判断するための先行技術調査に使われ、また他人の特許を潰すために公知例を探すことにも使われる。しかし最も価値のある使われ方は、研究者がその新情報を見て触発されてさらに次の技術開発に役立てるものであろう。

日本の特許情報は数年前までは毎年 35 万件程度、最近では諸事情で出願が減っており 26 万件程度が公開されている。10 年分で約 300 万件、20 年分で 600 万件もの詳細かつ膨大な技術情報が公開されている。特許情報は量的に膨大であるために、自分の欲しいものだけを探す事が難しい。そのためデータベースにおいては豊富な検索機能と参照機能が開発されている。検索の方法としては検索式とか論理検索といわれる and, or の演算子を使う方式が広く使われている。しかし、検索式作成を上手に行なう方法は習得が難しく慣れも必要であり、一般技術者が正確に利用するのはなかなか難しい。

それに対して、自分の必要としている技術内容を説明する文章そのもので検索できる「概念検索」の上手な使い方が徐々に明らかになってきている。本稿では、これまでに書かれた概念検索の使い方に関する論文を時系列で紹介す

る。

2. 論文の抄録

以下に抄録を示す。なお各抄録において「・・・と論じている」などの表現は冗長となるので避けた。また読者の理解のために、本稿の筆者以外が執筆しているものには簡単なコメントを加えた。文中、筆者とは本稿の筆者であり、元の論文の著者ではない。

2.1 特許情報を対象とした概念検索の試行

1998 年, INFOSTA シンポジウム, 清水美津子, 他, 情報の科学と技術, 48 巻 10 号(1998)

Concept Base Search を使った概念検索の有効性を「光触媒」などの短い質問文 2 例, 「血液成分を対象としたガンの検診に有用な検査・測定方法」の短文と一部変更した短文の 6 例で検証した。検索式による検索を概念検索に今すぐ完全に置換できるとは言えない, 予備検索用として利用できる可能性がある。

コメント・・・利用者の検索実験を含む最初の論文であるが, 約 1 頁の梗概しか入手できず詳細が分からない。

2.2 ビジネスモデル特許検索にも有効な概念検索

市川伸治, 野村総研, 「知的資産創造」2001.1 月号

「NRI サイバーパテントデスク」の概念検索でビジネスモデル特許の検索をおこなうという形で仕組みを解説した。単語や文章の意味情報を 300 次元の「文書ベクトル」で表し, あらかじめ数百万件の個々の特許の文書ベクトルを作成し, 入力文章(質問文)の文書ベクトルとの内積で類似度を算出し文章の意味により検索できる。検索式を作らず文章で検索でき, 特許番号指定で類似特許を探せ, 同義語や異表記の整備が不要である。同義語に関して「自動的に辞書を作成する」もので, 「カメラ」と「写真機」は意味の近いものと判断される。

コメント・・・「文章の意味」と書いているが文法的な処理はしていないはずであり, 単語の出現頻度の処理である

*むぐるま まさみち 六車技術士事務所
〒316-0032 茨城県日立市西成沢町 4-37-6
Tel. 050-8012-2416, または 0294-37-1284

(原稿受領 2014.11.29)

うが明確には書いてない。同義語に関して「自動的に辞書が作成される」と書いてあるが、(2003年頃でも現在でも同じであるが)同社システムを使ってこの論文でも例示している「カメラ」や「写真機」で概念検索して上位100件を見るとそれぞれかなり違う特許が出てくる。

2.3 特許情報検索の課題と概念検索システムの役割

六車正道, 日立製作所, 「知財管理」, 2001.12月

概念検索システムの原理として、質問文や蓄積特許でのワード(特徴ターム)の数(出現頻度)やワードの蓄積全特許中での出現特許件数の逆数を使ったものであることを説明し、関連度の評価ロジックを説明した。日立特許情報検索システムの使用経験から、上手な使い方のために質問文として「概念を明確に特定できる短い文章が最良」、不適切な質問文としては「短か過ぎて漠然とした概念の指定」、また、クレームや要約はその中の多くのワードを含む概念になるので不適切と分った。何度か繰り返し検索することの重要性、検索対象としては明細書が適していることを指摘した。調査専門家が時間をかけて行なった調査と数十分の概念検索の結果を比較し、情報検索に不慣れな利用者でも概念検索では高再現率を期待できる。また、概念検索の漏れが多いと指摘するよりも短時間で検索できる長所を生かすことが重要、アイデア育成支援への利用や、調査専門家が予備調査的にまた特許出願前の調査などで利用できる。

2.4 特許情報検索における全文検索と概念検索の役割分担

市川伸治, NRI サイバーパテント, 「特技懇」, 2002.5月, No.223

「NRI サイバーパテントデスク」の概念検索の仕組みと特長を紹介した。上手な文章入力(質問)法として、抽出したい主旨を的確に捉えた文章が指定できればよい、文章の長さによる検索精度の違いはない、文章の概念が焦点ボケになってしまうと期待した結果は得られなくなる。また上手く検索できない文章入力例として、否定形を含む文章、数字の範囲指定、文章への出願人や発明者の指定などがある。概念検索の役割として、出願前調査、アイデアを模索する局面での利用、異議申し立てや無効審判等の証拠資料調査で補完的に概念検索を使うことがある。

コメント・・・質問文の長さによる検索精度の違いはない、としているが同社および他社の概念検索システムを使った筆者の実験では大きく異なる。

2.5 企業における特許情報の活用

六車正道, 日立製作所, 「特技懇」, 2002.5月, No.223

検索対象としては要約や請求範囲よりも明細書全文の方が多くの場合検索結果が良好である。概念検索の質問文は核心をついた短文が適しているので専門サーチャよりも技術内容に精通した技術者に適している。要約や請求範囲をそのまま質問文に使うのでは多くの場合上手に検索できな

い、これは関係の少ないワードが多くなることで目標とする概念がボケるためと思われる。補助機能としてキーワードや出願人、IPC などによる絞り込み検索を併用すると再現率が大きく向上する。また、回答リストでは最重要の特許が先頭に来ないこともあるので先頭だけでなく50件程度まで見るのがよい。実際の特許出願前調査で、①調査専門員が約10時間で手めぐりと検索式利用で調査したものと、②特許技術者が約1時間で概念検索した結果を比較し、①:②=70件:50件であり、概念検索だけで30件を見つけた。概念検索によるアイデア発想として、具体的なアイデアがある場合のブラシアップと、課題をもとにヒラメキを期待する使い方を説明した。概念検索のようなインテリジェントな検索システムは利用法を元にした改善策を開発部門に還元することが重要である。

2.6 特許戦略を担う概念検索

星乃隆治, NRI サイバーパテント, 「NRI News」, 2003年3月

特許出願前の先行特許調査において特許検索に詳しくない研究者でも概念検索は有効である、概念検索は利用者が入力した文章の意味をシステムが解析して類似内容の情報を検索する。仕組みとしては文章から単語を抽出しその単語の意味を数学的に解析して数値化する。またある特許を指定するとその特許の文章を使って類似性の高い特許を検索でき、利用者は検索のための文章を作る手間を省略できる、文章での検索とこの方式の組合せでより目的に近い特許を抽出できる。また、技術分野ごとに概念検索を行なうことで検索精度を高められる。

コメント・・・「単語の意味を数学的に解析して数値化」は、例えば「半導体」が条件により導体や絶縁体になる物質ということを数学的に解析していると考えられるかもしれないが、キーワードの出てくる回数やキーワードを使っている特許件数などを指しているものと思われる。また、「技術分野ごとに概念検索を行なう」の詳細説明はないが、IPCによる限定を伴った概念検索のことだろうと思われる。

2.7 概念検索による特許情報の活用法 その1

六車正道, 日立製作所, 「発明」, Vol.100, No.4, 2003

利用法として、上位50~100番あたりまでに自分の欲しいものが存在しない場合には別の質問文を考えて再度検索する、それを3,4回繰り返すのが良い。また、概念検索の質問文に同義語の追加はほとんど不要ということを実例で示した。これは質問文中の他のワードが意味的に補ってくれることが期待できるからである。また、特定の同義語だけを入れるとそれらに偏った誤った検索になる恐れがある。質問は本質的には文章でも(複数の)キーワードでも同じであるが、キーワードでは不要な偏りを招くことがあるので文章が好ましい。概念検索は、検索式のような件数の絞込みは不要、質問文が絞込み過ぎた表現でも回答件数がゼロになることはない、対象とする技術概念を細かに絞

り込んだ検索に適しており広がりのある技術の全特許を検索するような用途には適していない、概念検索の操作法は簡単である、しかし質問文の作成は対象技術の知識が豊富であることが必要でエンドユーザーに適している。

2.8 概念検索による特許情報の活用法 その2

六車正道, 日立製作所, 「発明」, Vol.100, No.5, 2003

質問文の長さによる再現率の比較を実例で解説した。クレームや要約を質問文にすると 2,3 割の再現率であり実用にならない、数行の短文ではこれより上手く検索できることが多い。質問文が要約や明細書などの長い文章では上手くいかないことが多く、クレームでは「前記〇〇・・・」など発明の本質と関係の少ないワードを何度も使うことが多いので質問文としては不相当である。特許番号指定の概念検索もこれらの使い方になるために上手くいかないことが多い。最近の改良事項として、IPC やキーワードでの絞り込み検索の併用、またワードの配点の変更などを説明した。なお、絞り込み検索は概念検索の対象を絞ることであり利用者の工夫を反映できるので貴重である、絞り込み検索をやり過ぎると漏れが多くなるので多重の and は避けた方がよい。上手な概念検索として(1)質問文は短文が良い、(2)結果をみて数回概念検索をやり直すのがよい、(3)特許明細書に使われているワードを使った方がよいが徹底しなくてもよい、など 7 点をあげている。

2.9 概念検索による特許情報の活用法(アイデア発想支援ツールとしての活用)

六車正道, 日立製作所, 「発明」, Vol.100, No.6, 2003

概念検索はアイデアの創生、発明の保護、権利の活用という特許サイクルの全てにおいて役立つ。エンドユーザーは、概念検索に必要な質問文は既に知っており好適な検索システムである。概念検索をアイデア発想支援として利用するやり方を具体的に示した。また、アイデアや課題と同じ技術を探すだけでなく意図的に他の分野を探すことも有益であると実例で示した。漏れのある検索でも役立つとし、固定料金制データベースの優位性、また実際の研究者へのアンケートを紹介した。さらに、特許出願前の調査では 1,2 時間程度の調査でも一定の成果をえられること、権利活用場面でも有益であり、特に明細書の片隅に書いてある事項の検索も可能であり、特許係争対策用にも有益である。

2.10 特許情報における主題情報の検索：概念検索とその限界

高橋昭公, (有)テル・リサーチ, 「情報の科学と技術」Vol.54 (2004), No.7

概念検索システムでは、形態素解析、重み付け、類似比較の 3 要素が用いられ、類似比較では一般的にベクトル空間モデルが採用されている。NRI サイバーパテントデスク、PATOLIS-IV、G-Search/ATMS の 3 システムを比較した。権利関係の調査にも適用できるかどうか特許番号指定やクレーム、要約を種文書として検索実験を行い 4 件の

公報引用特許のうち 1 件を抽出できた、IPC の絞り込み検索で引用特許を上位にランクさせられた、概念検索でもシステムの理解など検索スキルが重要でその習熟には数多くの実務経験が必要である。特許情報における概念検索システムの特色を整理し、調査目的別に概念検索をどのように使い得るか説明し、概念検索の限界も論じた。

コメント・・・検索スキルの 1 つに「質問文章の最適な作成方法」を挙げているが、検索実験では質問文としてクレームや要約を使っている。「概念検索の利用の習熟には数多くの実務経験が必要」という部分は筆者は賛成しかねる。

2.11 概念検索システムの現状と使いこなしの検討 — 知財強化に貢献する概念検索 —

六車正道, 日立公共システムエンジニアリング(株), 発明, Vol.102, No.4, No.5, 2005

4 月号; NRI サイバーパテントデスク, 日立総合特許情報検索システム, PATOLIS, PATOLIS-J, ATMS/G-Search, RIPWAY/リコーテクノシステムズの機能、性能を比較した。性能は、生ゴミ処理脱臭とチタニア超薄膜の 2 テーマの検索実例で比較した。質問文を利用者作成の短文、第 1 請求範囲、要約、明細書で、また検索対象を明細書、請求範囲、要約のテーブルで比較し、最良の組合せとして質問文は短文、対象は明細書であることを明らかにした。なぜ請求範囲を質問文としたのではうまく検索できないのか、短文のワードと比較検討した。さらに、稀に請求範囲の文章が質問文に適している場合がある実例を示した。また、1 つの短文だけでなく質問文を変えて複数回行うことで再現率が高くなることを実例で示した。特許番号指定の概念検索では複数特許を指定することで再現率が高まる場合と下がる場合を示した。

5 月号; 多くの概念検索システムでは「モータ」で検索すると、モータの代わりに「電動機」としか書かれていない特許は検索できない。しかし概念検索は 1 語での検索は妥当な利用法ではなく、複数のワードやそれらを含む文章で質問すべきである。質問を複数のワードだけで行なう場合より文章で行なう方がよいと実例で説明し、再現率を高めるためにテーマを分割して質問する方法や、出願人、IPC、キーワード、発行期間などで検索対象を絞り込んで概念検索する方法を説明した。概念検索が知財部門、研究部門で役立つ場面を整理して示した。特許調査のステップフローを示して概念検索の役割を説明した。

2.12 概念検索の現状 — 概念検索システムの比較と使いこなし —

六車正道, 日立公共システムエンジニアリング(株), 「JAPIO/創立 20 周年記念誌」, 2005.10 月

概念検索の原理を示し、システムを比較し、現状を次のようにまとめている。①再現率は 1 割以下のものから 8 割を越すものまで様々のものがある。②概念検索で審査官引例のような権利的な調査もカバーしうる。③概念検索は実用になるものがすでに存在するとみてよい。概念検索の上

手な使い方の1つとして、質問文として技術的説明文に加えて目的や効果などを入れることを述べた。

2.13 概念検索システムの有用性の研究と提言

知的財産情報検索委員会第2小委員会, 知財管理, 55巻(2005年), 11号

概念検索の仕組みと特徴を, ATMS/ジーサーチ, NRIサイバーパテントデスク, PATOLIS-IV, RIPWAY/リコー, Sharesearch/日立で比較した。また, 課題と適合特許リストを整理したテストコレクションを使って34課題で検索実験を行なった。質問文はほとんどが第1請求項を使い, 検索対象はほとんどで請求項と明細書全文の2つとし, 上位500件までに出了かどうかで判定した。なお, 部分的には100位や50位までの一致も検討した。結果として検索対象は請求項よりも全文が良いとし, 適合特許の請求項数の違いによる影響の検討, IPCセクションを使って技術分野による影響の検討などをした。例えば, 課題(質問)特許でディスプレイと書かれている場合, 適合(回答)特許で表示装置と書かれているものはランク外になったものがあり, 単独で潰せるA判定特許が上位50位までに出るケースがシステムによるが14~24課題あった。

コメント・・・数多い検索実験の結果でありその価値は大きい。ただし, 500件まで比較しているがこれは500件まで目視チェックする使い方を想定しているのであろうか。また質問文を請求項とするのは概念検索の利用法として最上であろうか。

2.14 概念検索の活用法 — 概念検索システムの特徴と知財実務への応用 —

六車正道, 六車技術士事務所, 「Japio2006YearBook」, 2006.11月

概念検索の特徴として(1)自然文で質問できる, (2)操作が簡単に行える, (3)ヒットの可能性の高い順にリストアップされる, (4)詳細技術に絞り込む検索に適している, (5)依頼者に検索条件を説明し難い, をあげた。特長を生かした利用として(1)研究テーマを探す場合, (2)研究を深めていくときの相談相手としての利用, (3)特許出願前の調査, (4)製品出荷前などの侵害予防調査, (5)無効資料調査を上げ, 概念検索を含めた特許調査フローを図示した。さらにアイデア発想支援として概念検索が適している理由として, 考えている内容をそのまま質問文に使うことができ, 検索式作成など手法にとらわれて思考が中断することがなく連続できることがメリットである。また, 概念検索を利用したアイデア発想支援の実際のやり方の基本と実例として「迷惑電子メール対策」を示した。

2.15 概念検索についての誤解

六車正道, 六車技術士事務所, 「Japio 2007YearBook」, 2007.11月

概念検索への誤解と思われる事項を整理した。誤解を生じやすい原因として, ①重み付けをおこなうこと, ②操作

法を利用法全体と考えてしまうこと, ③過度の期待とその反動, とした。まず(1)何をしているか明らかでない(分からない)という誤解に対し原理を説明した。次に(2)役に立たないとする誤解として, 概念検索の再現率は低い, 本格調査の手がかりを得るための予備調査に使うもの, 大雑把な検索に適しており絞り込んだ検索には適していない, 技術的調査には使えるが権利的調査には使えない, 同義語や類義語の入力は重要である, 上手な質問文の作り方や概念検索のやり方は検索式と同じくらい難しい, 利用者による工夫の余地は少ない, その他を上げて誤解であることを説明した。さらに(3)何でもできるとする誤解として, 思いついた2,3のワード入力が必要な特許を全部探せる, 質問文は要約や請求範囲または明細書でよい, 1つの質問だけで欲しい特許を全部探せる, 回答リストの第1位の特許が最も関連の高いものである, 概念検索の利用法や検索目的の技術内容を知らない人でも簡単に検索できる, 短時間で高再現率が期待できる, 概念検索だけですべての検索が行なえる, 動向調査が簡単に行なえる, 現在の概念検索は完成しておりしばらく改良されることはない, をあげてまちがいを示した。

2.16 特許文書構成と分類情報を用いた類似特許検索方式の精度評価

間瀬久雄, 日立製作所, 「Japio2007YearBook」, 2007.11月

TF-IDFの原理による類似特許検索を, 350万件を対象に正解を審査官の拒絶理由の引用特許とし, 数千件の実験テーマ数でおこなっている。質問文は元になる特許の請求範囲, 明細書特許, 要約などを使い, 元の特許の特許分類と組合せている。あるキーワードが請求範囲, 要約, 明細書の全部に出てくればその重みを大きくした。また, 元の特許と同じ特許分類をもつ特許は類似度スコアを上げる。このようにシステムをチューニングすることで再現率が向上した。

コメント・・・システム開発の改善であり, 利用上の工夫ではない。

2.17 概念検索はなぜ上手に検索できるのか?

六車正道, 六車技術士事務所, 「Japio2008YearBook」, 2008.11月

要約や明細書がなぜ質問文として多くの場合不相当であるのか検討した。要約と同じ長さでキーワード分布は短文に近い文章を作成して質問文とした場合, 上手な概念検索が行なえた。このことから文章の長短が問題ではなく「キーワード分布」が目的とする技術内容に近いかが重要と述べた。要約や明細書では文字数が多いだけではなく, キーワード分布が異なってしまう実例を示した。上手な利用のための工夫として, 技術を説明する具体的なキーワードで40~80文字程度の文章とし, 絞り込んだ技術の検索に使う, 同義語を使いたい場合は別の質問文としてやり直す, 上位500件程度までは目視チェックすることとした。

2.18 概念検索技術および特許検索への適用可能性について

八木敬宏, 間瀬久雄, 岩山真, 日立製作所, 「特技懇」, 2009.1.30, No.252

概念検索の原理は、ベクトル空間モデルと確率モデルさらには言語モデルも単語（特徴語）の出現頻度を使うことでは大差がない。特許検索への適用として、本願の理解のための情報収集、検索式作成支援、検索式検索の結果を概念検索で並べ替えて参照するスクリーニングへの適用、分類付与支援で期待できる。また課題として特徴語の抽出や重み付け、類似度計算に加えて特許の文書構造、属性情報、引用関係の利用などをあげた。さらに今後の検討課題として、技術分野別のアプローチとか、本願全体を入力（質問）文章とするのではなく今どこに着目しているか指定するようなユーザの視点や意図が検索に反映される仕掛けを持つこと、結果を見ながら概念検索を繰り返す仕掛けが精度向上になる。

コメント・・・「今どこに着目しているか指定する」「検索を繰り返す」ための仕掛けが必要とすれば、それが無い現状では利用者が技術を説明する短文を作り、その概念検索の結果を見てさらに変更して複数回の検索を行うことが上手な使い方と理解される。

2.19 特許審査業務への適用に向けた概念検索精度の評価 間瀬久雄 日立製作所 「JAPIO 2009YearBook」, 2009.11 月

特許庁の審査関連業務に概念検索を利用する場合、類似内容の複数の出願案件を集める段階、本願の出願人が過去に出願した類似文献の収集、先行技術調査、検索結果の優先順並べ替えなどに利用できる。検索精度向上に関しては、特徴語の抽出、重み付け、類似度計算などをあげている。審査用の概念検索のプロトタイプや結果のチューニングの紹介をした。

コメント・・・精度向上に言及あるがシステムの改善策の検討であり、現存システムの使い方の具体策はほとんど示されていない。

2.20 特許を対象とした概念検索の技術課題

間瀬久雄 日立製作所 「JapioYearBook2010」, 2010.11 月

請求範囲からでなく全文から抽出した重みの高い特徴語で概念検索を行う方が精度が高い。さらに精度が低下するのは、発明概念の特徴を考慮せずに特徴語を抽出しているためである。また特徴語の同義語展開は厳選して行えば効果的、近傍語を意識した特徴語抽出で精度を上げられるだろう。

コメント・・・特徴語抽出に目標の概念を考慮しない概念検索では再現率が低いとのべている。当然である。

2.21 概念検索における質問文の長さに関する考察

六車正道, 六車技術士事務所, 「JapioYearBook2010」,

2010.11 月

概念検索で再現率を高めるには、調査したい技術の目標の絞り込みと検索の絞り込みの程度が一致すること、つまり「キーワード分布」の一致が必要である。概念検索では 10 文字程度の短すぎる質問文では多くの場合必要とされる詳細な技術概念に絞り込むことができなく、逆に 200 文字程度以上の長い質問文や特許番号指定の概念検索では不要なワードの増加によりの定まらない質問となり、共に再現率の低いことが多い。しかし、短すぎる、または長すぎる質問文による概念検索は利用が簡単である大きなメリットがある。そこで、予備検索的に利用するとか、必須の IPC やキーワードによる絞り込み検索の併用、検索式による検索との組み合わせなどにより概念検索を上手に利用できる。

2.22 特許概念検索における特徴語抽出に関する評価と考察

間瀬久雄, 日立製作所, 「JapioYearBook2011」, 2011.11 月

クエリ（目標特許）から統計的に抽出した特徴語だけの概念検索よりも、意味的解析によって抽出した発明概念特徴語を加えた概念検索は再現率が高い。統計的な特徴語の抽出は目標特許の全文から重みの高い 10 語を選定し、発明概念特徴語の抽出は目標特許の要約を読んで内容を理解して設定した。

コメント・・・「目標を理解して質問文を作る」ことの大切さに気付いたということは大きな前進であり今後の研究の方向に期待する。

2.23 特許情報を概念検索で利用するアイデア発想支援

六車正道, 六車技術士事務所, 「JapioYearBook2011」, 2011.11 月

KJ 法とか TRIZ などこれまでの発想法は手間がかかり実用的ではない。それに対して、解決すべき課題に関連する情報を学んでおけば、解決したいという強い目的意識がある場合、アイデア発想は無意識的に生じる。関連情報を得るやり方として、特許情報を概念検索で利用することを勧め、概念検索の上手な利用法を説明した。

2.24 特許概念検索結果の理解支援に関する考察

間瀬久雄, 日立製作所, 「JapioYearBook2012」, 2012.11 月

概念検索結果の多くの特許内容のマクロな理解支援のために、特徴語一文書マトリクスを使ったものを紹介した。このマトリクスは縦軸に入力文章から抽出される特徴語を並べ、横軸に検索結果の上位の特許を並べた表である。特徴語の検索結果文書における重要度を濃い色でぬりつぶして、視認性を高めた。

コメント・・・現存するシステムの紹介ではなく試作的、実験的な報告である。

2.25 概念検索を利用した無効資料調査

六車正道, 六車技術士事務所, 「JapioYearBook2012」, 2012.11月

概念検索は絞り込んだ検索に適しており, これを使って無効資料調査の「コア検索」を行なう。その後, 信頼性保証のために検索式で検索した大量の特許を目視チェックする。「手続き代行システム」というテーマで, 概念検索で無効資料調査を行なった調査例を紹介し, 概念検索を利用するメリットを説明した。

2.26 単行本「技術者のためのアイデア発想支援 ― 特許情報を概念検索で使いこなす」

六車正道 (六車技術士事務所), 『発明推進協会』, 2013.5月

特許データベースを概念検索で利用することで, 利用者の必要としている特許を容易に入手できる。それを読むことで頭脳が刺激されアイデア発想支援が可能になる。事例3件あり。概念検索の上手な利用法として, 質問文は利用者が探したいと考えている技術を説明する短文を使うこと, 上位50件くらいまで見ること, 質問文を変えて数回検索することなどを上げている。

2.27 類似画像検索と概念検索を統合した特許検索システムの構築

秋山直人, 岩山真, 日立製作所, 「JapioYearBook2013」, 2013.11月

特許の文章による概念検索と図面による類似画像検索を組合せたメディア統合検索を, 10年間分の特許を対象に, 検索要求(目標)特許に類似の特許を検索した。質問文章に請求範囲を使った場合より類似図面のあるものに絞ることで適合率が高くなった。図面は全図面より選択図面の方が適合率が高い。

コメント・・・全図面よりも選択図面が良いことは, 質問文は検索者が作成した文章が適していることを示唆している。

2.28 概念検索の活用促進7つの関門

六車正道, 六車技術士事務所, 「JapioYearBook2014」, 2014.11月

概念検索の活用促進に対して立ちはだかる問題を以下の7つにまとめ, 原因と対策を示した。(1)システムによる再現率の大きな違い, (2)提供側の指導不足, (3)上手な使い方が分からない, (4)再現率が低いという誤解, (5)適用業務が分からない, (6)利用価値なしと思いつむ(上手な使い方を知らずに使った経験を元に), (7)依頼元が検索式を求める(その結果, 調査員が概念検索を敬遠する)。

3. 考察

3.1 概念検索実験の数と質

前章で紹介した記事に出てくる検索実験には2つの系統がある。1つは数多くの実験を行なうもの, 2つ目は実験

の数は少ないが綿密な検討を行うものである。

(1)数多い比較検索実験

このやり方では質問文は請求範囲や要約などを使う。回答目標としては審査官引例などを利用するものが多い。数多く検索実験することによる知見を得るメリットがあろうが, 質問文が概念検索の特徴に合ったものであるかという観点の吟味は行えない。また目標とする回答が最良のものであるかどうか不明である。

(2)実験数が少ないが綿密な検討

この場合の質問文は, 検索者が目的の技術内容を熟知して作成した, つまり意志を反映した短文である。回答目標はいくつかの方法で調査した結果の合計を用いることが多い。実験数が少ないことによる偏りがありうるが, 概念検索に適した質問文の検討や, 最適と思われる質問文を使つての各システムの相違を論じるのに適している。

論文2.20は(1)のやり方であるが, 精度が低いのは検索目的の技術内容を理解せずに特徴語(質問キーワード)を選定する場合であると気付いている。また2.22も(1)のやり方だが, この論文では検索者が目標特許の要約を読んで意味を理解して特徴語を加えたら再現率が高かったという。これらの意味する所は, 検索目的の技術内容を理解して特徴タームを選定すれば精度が上がるということであり(2)のやり方に通じるものである。二つのやり方の接近は好ましいことであろう。

3.2 何を検索するのか?

質問文として請求範囲や要約, 極端な例としては明細書全体を与えて行なう概念検索は一体何を検索しているのだろうか?, 特許番号指定で概念検索を行なう場合もその特許の請求範囲や要約のうちの何を質問文に使うか宣言する(または決まっている)ので結局同じ疑問が生じる。利用者は文献の全体が同じ物を本当に探したいのだろうか?

この疑問を提示する理由は, 通常の利用者は文献の全体が同じでなくても一部分に自分の期待することが書いてあるものを探していると考えからである。

概念検索は質問文と同じキーワード分布になるものを探すシステムである。キーワード分布とは前章の記事2.17, 2.21にも書かれているが, 検索に使うものとして質問文からシステムが自動抽出したキーワード, および出現頻度と使用文献数である。質問文が質問したい事項を正確に表現するキーワード分布になっておれば, 概念検索の結果は再現率が高くなる。

概念検索を行なう場合, 利用者はいかなる技術も目標として実施しているが, 請求範囲などを質問文とすることに対して2つの疑問がある。1つは, 利用者の目標技術がその請求範囲と全く同じかどうかである。先行技術調査などではそのような事は少ないであろうが, その請求範囲を潰す無効資料調査では同じといえる。また, 検索実験をするのであれば同じと言ってよからう。

2つ目の疑問は, 請求範囲のキーワード分布が探したい技術のキーワード分布になっているかという問題である。

これは概念検索の根幹にかかわる問題である。請求範囲は人間が文法を頼りに技術内容を理解するものであるが、概念検索は文法を使わず出現頻度と使用文献数により類似度を評価するものである。例えて言うと、単語の多寡で判断する赤ん坊や幼児に対して、文法で理解する大人の言葉で語りかけて理解できるか、というようなことである。つまり、無効資料調査などのようにその請求範囲と同じ技術内容の文献(特許)を探したい場合でも、その請求範囲のキーワード分布が目的どおりになっている保証はない。というよりも、「前記○○の・・・前記○○の・・・」などのように重要でない言葉が頻繁に繰り返して書かれる多くの請求範囲においては不都合なキーワード分布になる恐れの高くないことが予想される。要約や抄録は請求範囲ほどではないが類似の問題がある。

このように考えると請求範囲や特許番号指定での検索実験はシステム開発時とかごく参考的に行なう検索として位置付けるべきものであり、概念検索を知財関連業務に使うために再現率高く使いこなそうとする議論にはあまり役に立たないやり方のように思える。

4. 終わりに

例えばピアノにおいて、鍵盤の滑らかな弾き易さは製造者の責任であるが、上手な弾き方に関しては多くの利用者の工夫の蓄積が大きな要素であるに違いない。概念検索もシステム構築者の責任事項とは別に、利用者の工夫に負う事柄があるのではないか。

紹介した記事において、本稿の筆者以外が書いているほ

とんどのケースで質問文を請求範囲とか機械的に抽出したものを使っていることにあらためて驚いた。概念検索の性能比較とか上手な利用法の検討に、利用者の検索目的を反映しにくいこのようなものをなぜ質問文に使うのであろうか。たぶん質問文の標準化のためであろうが、もう少し工夫をすべきである。

取り上げた記事はすべて過去のものであり、すでに新しい試みをしている利用者もいるであろう。新しい試みが公表され、工夫の蓄積となっていくことを期待する。

将来、2014年頃概念検索に関する論文を振り返ってみれば、目の見えない人が象のあちこちを触って「大きな団扇(うちわ)のようなものだ」「いやいや、大きな木のようなものだ」「それらは全部間違いで、穴の開いたホースのようなものだ」と言いあっているような状況かもしれない。たぶん我々は、概念検索というまだ幼稚ではあるがインテリジェントな検索システムの評価基準や使いこなし法、またそれらの表現法を十分に知らないのだと思われる。

本稿でまとめた抄録が原本への興味をかきたて、ひいては概念検索への認識を深め、結果として上手な使いこなしが進み、システムの改善に役立つことがあれば望外の喜びである。

本誌では通常、論文末尾に参考文献を記すが、当原稿では本文中の各節にその情報を記載した。なお、下記 URL に各論文への可能な限りのリンクを張った資料がある。

<http://www.patentcity.jp/patentcity/gainen.htm>

Special report: Study articles concerning useful use of concept search method. Masamichi Muguruma (Muguruma Professional Engineer Office, 316-0032 4-37-6, Nishinarusawa-cho, Hitachi-shi, Ibaraki-ken)

Abstract: Many concept search systems have equipped in Japanes patent databases. The number of study articles concerning useful use of concept search method is nearly thirty. The other hand, there is some opinion that the concept search is not useful system. Author wants to make clear the problems and tried to show abstracts of those study articles. There is two types of search experiment. First one is many-numbers theme experiment and the other is a-few-numbers. As search request sentence, patent claim is used on the many-numbers type experiment and user-made short document is used on the a-few-numbers type. By those abstracts, it is considered that some differences make clear between a high-rate-recall concept search and a low-rate.

Keywords: concept search / patent search / information retrieval / database / key word distribution / appearance frequency