

概念検索は役に立たないのか？

Is a concept search useless?

六車技術士事務所 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）。

✉ メールフォーム： "mailto:mug" [検索] ☎ 050-8012-2416

1 はじめに

昨年末までに公表された概念検索の使いこなしに関する論文を集めた記事がある¹⁾。これを見ると概念検索の利用に関する記事は筆者の執筆したものが非常に多いことが分かる。筆者は、現存する概念検索システムは実用になる程度の再現率をすでに備えているものがあり、それらは比較的手軽に利用できるものなので研究開発者や特許調査員を含む多くの人に利用されるべきであると考えている。しかしこの考えは、特許情報の関係者は必ずしも受け入れていないようである。概念検索は本当のところ役に立たないのだろうか？

そこで今回は、なぜ筆者は概念検索に対してこのように旗をふって応援しているのか明らかにすることを通して、概念検索を理解する一助になることを計画してみた。

2 生物と機械のやり方の違い

生物のやっていることと類似のことを機械が実現する場合、似ている点と異なる点がある。よく言われる例であるが、飛行機は羽根の断面の形は鳥を真似したが、羽ばたくやり方は真似せず固定翼にすることで成功した。羽根を羽ばたくことまで過度に真似した多くの人の試みは全て失敗に終わったのである。

人は複雑な論理を扱うのは上手だが、単純労働の繰り返しは下手であり間違ったり怪我をしたりする。一方、機械は逆で単純作業の繰り返しほど上手であり、特にソ

フトウエアでは繰り返し回数がどれほど多くても全く損耗しない。人は 10 種類の要素を使う 10 進法を使って 7,8 桁程度までの取り扱いが得意である。コンピュータは 2 つの要素を使う単純な 2 進法を使い 32 桁とか 64 桁という膨大な計算を楽に行なうことができる。そして現在のコンピュータ社会は、基礎部分でコンピュータに 2 進法による計算をさせその結果を人間の得意な 10 進法として表示させることで成立している。

文書の類似度の比較においても人とコンピュータはそれぞれ得意の手法を利用するべきであろう。つまり、人はキーワード間の関係を表す複雑な文法を使うが、コンピュータによる概念検索はキーワードの出現回数を数え上げる方法を用いる。概念検索が文法を扱わないという理由で劣っているという考えは、過度に人間のやり方を真似させようとする偏った考えであると筆者は考えている。

ところで、キーワードの出現回数を数えることで「ぼっちゃん」の面白みが理解できるだろうか。出来るはずがない。それは目的が違う。

情報検索が求めているのは文章の面白みの理解ではなく、文章が似ているかどうかである。固定翼にしたことで小鳥のように優雅に空を飛ぶことはできないが、驚くほど遠方まで大量の物資を高速で運ぶことができる。特許情報も概念検索による利用を進めることで、文章の類似性については大量で迅速に活用できるようになるはずである。

3 重み付けは知的な活動

検索式は、質問として与えたキーワードの有無だけで判断するものであり、論理は明快である。したがって、機械の動きは全て理解したいと考える多くの利用者（探索者）は頼りにし易い。しかし、単純であるために再現率を上げるには多くのノウハウを必要とし、上手に利用するのは容易ではない。

これに対し、概念検索はキーワードの有無に加えてそれに重み付けする、つまり重要度を取り扱うので検索式よりも知的と言えよう。その結果、利用者の意図をくみ取ったような検索結果を比較的容易に得ることができる。

3.1 概念検索とは何か？

「概念検索」は質問として利用者が与えた「短文」の中からキーワードを抽出し、それに「重み付け」を「システムが自動的に」行ない、それに基づいて対象特許の「類似度の計算」を行い、得点の高い方から表示するものである。ところがこれを納得できないと言う人がいる。概念検索を使いたがらない人は、上の「 」を付けた事項に違和感を持つようである。

以下にその違和感をQとし、回答をAとして纏めてみた。

Q：「概念検索」といかに心があるような言い方をするのはおかしい。

A：我々は鳥も飛行機も「飛ぶ」と言う。世の中では「コンピュータの判断は・・・」などの言い方もされるが人のように心で判断している訳ではない。一方、概念検索はキーワードの有無だけでなく、キーワードに重み付けをして利用者の意図（希望する技術概念）を察知したような結果を示すことができる。したがって、概念検索と言ってもそれほど間違いではないと考える。

Q：手持ち特許に近いものを探す場合は質問文を「短文」で作るなど不要のはずだ。

A：人に頼む場合、手持ちの特許のどの部分の似ているものが欲しいか伝えれば焦点の定まった良い結果を得ることができる。検索式も概念検索も同じでポイントを絞る必要がある。概念検索はポイントを示す文章をそのま

ま質問文に使うということである。

Q：キーワードの「重み付け」は様々な観点があるのでコンピュータに出来るはずがない。

A：似ているかどうかの観点での重み付けである。そのため、質問文中と対象特許のキーワードの回数は多い方が重要で、データベース全体においては回数の少ないキーワードが（識別性が高いので）重要として重み付けられる。

Q：「システムが自動的に」行うのでは利用者の意志が入っていない。

A：質問文に利用者の意志を含めることができる。意志の入っていない適当な質問文を使っておいて、自分の欲しいものが出てこないというのは間違った使い方である。請求範囲や要約文や明細書全文を質問文に使ったり、特許番号指定などの概念検索は利用者の意志の少ない不明瞭な質問になっている。

Q：「類似」かどうかは専門家が行なうべきもので「計算」など出来るはずがない。

A：最終判断はコンピュータによるラフな評価だけでなく、技術や法律の専門家による判断が必要なことは当然である。検索式はキーワードの有無の判断であり、概念検索は有無だけでなく重み付けを加えた融通性のある判断をしているが、専門家の総合的な判断とはレベルが違う。

3.2 「知的」と言えるのか？

ところで、概念検索を含むコンピュータの活動を知的と言って良いのだろうか？

現代の理工学においては、コンピュータが利用者の意図を取り扱うとか、利用者の表立った行動の裏の思いを取り扱うことを知的と言っているようで、「賢い」という表現もされる。

ものごとの判断において0か1、有りか無しという2つに分けるのは最もシンプルであり、コンピュータは取り扱いが得意である。しかし人間は、単純な0と1ではなく、その間に膨大なグラデーションを認識することができ、その組合せにより知的活動が生じる。

音の有無だけの記録、再生ならばボツボツという音が聞こえるだけで音楽にならない。しかし、CD（コンパクトディスク）で情感のこもった音楽を記録、再生でき



るのは、システムが約 2 万分の 1 秒ごとに音の大きさを最大 1023 種類の大きさに分けて記録するからである。非常に短い時間ごとに大きさを数百の大きさに分けて記録する（重み付けする）ので、感情表現を盛り込んだ音楽として聞くことができる。

検索式ではキーワードの有無しか扱えない。しかし、概念検索では TF・IDF と言われる技術によりシステムがキーワードに重み付けを与える。筆者がその方法にそって計算してみたところ²⁾、ある質問文のキーワード中、TF・IDF 値が最大のものは 13.58 であり最小値は 0.075 であった。その差は約 180 倍となる。CD の場合と比較するのが妥当かどうか分からないが似た値である。

質問文中のキーワードが 10 個であれば 10 次元の座標において方向と大きさを持った値として質問が表現され、類似の値を持つ特許が回答の上位にランクされる。キーワードの有無だけで判断する検索式に比べて知的な検索が行なえる、と言って良いと考える。

3.3 重み付けの方法

TF・IDF によるキーワードの重み付け法を簡単に説明する。図 1 はその定義を示す式である。TF は Term Frequency の略であり、質問文の中で使われている各キーワードの回数（頻度）である。TF が大きいほど重要なキーワードとなる。

ところで、使用回数の多いキーワードが必ずしも重要ではない。しかしコンピュータに適した方法であるので採用されている便宜的な方法と考えればよい。利用者としてはそれを承知して質問文を作る必要がある。車で道路を走るにはルールがあり、それに従えば上手に運転できる。

なお、データベース中の各特許における TF はその特許の中に出てくる各キーワードの回数である。

また IDF は Inverse（または Inverted）Document Frequency であり、あるキーワードがデータベース全体において使われている特許件数を、そのデータベース全体の特許件数で割った値の逆数（の対数值）である。IDF が大きいほどデータベース全体で珍しいキーワードといえる。

筆者はこれまで質問文は短文が良いと主張してきた。これは、長文よりも短文の方が利用者の意志が反映され易いということである。ある特許の長い明細書を示してこれに近いものを探せというだけでは、長い明細書の中のどの部分に焦点を合わせれば良いのか明確でない。人に依頼する場合もそうだし、ましてやアタマの悪いコンピュータではなおさらである。350 文字程度の要約でも似たことがいえる。

これに対して、概ね 30 ～ 80 文字程度の短文にすることで多くの場合、利用者の意志を明瞭に絞り込める、つまり知的な質問文になる。

ただし、この数字は絶対的なものではない。つまり、必要なキーワードが適切に使われているものであればこれよりも長い文章でも構わない。以下、複数のキーワードとその得点を「キーワード分布」という。質問文のキーワード分布が検索目的と一致していれば、長い質問文でも構わない³⁾。ではなぜ筆者は「質問文として短文が適している」と主張しているかと言えば、短文にすることで不要なキーワードの入る余地が少なくなり、キーワード分布が目的とずれない質問文、つまり良い質問文になることが多いからである。

4 知的プロセスは説明し難い

概念検索は、キーワードの単純な有無でなく、知的な

$$TF \cdot IDF = \left(\text{キーワードの出現回数} \right) * \log \left[\frac{1}{\frac{\text{キーワードの使われている特許件数}}{\text{データベース全体の特許件数}}} \right]$$

図 1 TF・IDF の定義

システムであるがゆえに説明し難い、理解し難い点がある。

例えば、車のエンジンの基礎は誰でも理解できるが、コンピュータにより複雑に制御された詳細を理解することは多くの運転者には困難である。さらにブレーキも複雑な制御がなわれ、ハンドル操作も支援だけでなく自動化さえもなされようとしている。もし、複雑な制御を理解できないから事故が心配と考えるならばとても運転できないだろう。現実には多くの人が安心して運転しているのは、製造者への信頼と安全に使えている実績への信頼があるからであろう。

概念検索も製造者が信頼に足ることを研究し PR すべきだが、利用者も有益な使い方を探求すべきと考える。なぜなら、自動車ほど一般的なものではないから製造者のコスト負担も限定的だろうと考えても良いし、製造者だけでは分からない領域があると考えても良いし、そうした方が利用者自身のメリットになる、と考えても良からう。

TF・IDFの論理を一応は理解していても、複雑であるために自分が与えた質問文が妥当なのか判断し難く、また検索結果をどう扱えば良いのか分かり難いことがある。複雑になってしまうことは、知的であることの避けがたい宿命でもあるようだ。

(1) なぜその質問文にしたか？

前章でも説明したとおり概念検索の質問は短文が良いのだが、短文であるために文章中のキーワードが少し異なるだけで結果に大きな影響が出る。調査結果の良否に関わらず、質問文選択の理由を説明するのは容易ではない。例えば以下2つの質問文で考えてみよう。

質問文1「インターネット競り取引で購入希望価格と最大許容値を入力できる」

質問文2「インターネット競り取引で購入希望価格と最大許容値を入力できる自動取引システム」

質問文1で上位に妥当な特許が見つかり、2ではその順位がずっと下だった場合、なぜ質問文2では「自動取引システム」を付けるのが不都合なのか問われても合理的な説明が難しい。

あえて言えば次のようにならうと思われる。“質問文1の複数のキーワード群の意味する技術範囲には「自動取引システム」がすでに含まれているので、旨く検索

できる。一方、質問文2のように「自動取引システム」のキーワードを追加するとそれが強くなり過ぎて、「自動取引」に関する様々の特許が上位に来て、その結果「最大許容値」まで絞り込まれた特許の順位が下がる結果になる。”

一方、検索式の場合も選択したキーワードやそれらの論理関係の説明を求められるが、この場合は論理関係がand.orのようにシンプルであるために比較的説明し易い。

(2) なぜそこまで見たか？

ある検索式で回答が1000件を超えれば、もう少し減らすために多くの場合は追加の条件でand検索をして例えば150件まで減らしてその全件を目視チェックする。

しかし概念検索の場合は、類似度の高い順に回答されるものであるために、ランクの低いものを含めると該当件数は多くの場合100万件を超える。このため類似度の高い、つまりランクの上位何件までを目視するか利用者が決める必要がある。目視件数の正確な決定は過去の経験に基づくかなり高度な判断であろうが、経験が少なくても適当な件数で止めることはできる。いずれにしても、その合理的な説明は難しい。

目視件数の決定は、質問文の内容の妥当性と限定性の2つの観点で決めることができると考えている。内容の妥当性とは目標とする技術と質問文のずれである。限定性とは技術的な細部まで絞り込んであるかどうかということである。

実際の利用に際しては、次のような逐次推定方式とでもいうやり方で目視件数を決めることができる。

① 10（～20）件までにヒットあり

⇒ 50件近くまで見る

・ その後はヒットなし ⇒ 50件で終わり

⇒ 別の質問文を検討

・ 30件付近にヒットあり ⇒ 100件まで見る

・ 60件付近にヒットあり

⇒ その後なければ100件で終わり

・ 100件近くにヒットあり

⇒ 「質問文が絞られていない」

⇒ さらに150件目まで見る

② 10（～20）件までにヒットなし、近い内容の特許



はある

⇒ 50 件目まで見る

・ 30 ~ 50 件目にヒットがあった

⇒ 100 件目まで見る

・ 50 件目まで見てもヒットがない

⇒別の質問文を検討

③ 10 (~ 20) 件までに近い内容がない

「質問文がずれている」 ⇒別の質問文を検討

技術的な絞り込みの少ない質問文では、300 件目や 500 件目まで見ても頻度は低いが類似の特許が見つかることになる。そのようなワイドレンジの概念検索は検索式による検索と似ている。

概念検索が上手に使えた場合でも合理的な説明が難しい場合がある。これは、調査員に技術内容の詳細を指定する調査を依頼する場合と似ている。それは知的な調査依頼だからである。これに対して、あるキーワードを持っているとかある IPC (国際特許分類) の全件ピックアップなどの単純な依頼は内容の説明が簡単である。

5 データベース検索は大まかな絞り込み

あらゆる調査と同じく、特許調査も何かの手がかりを使って対象を大まかに絞り込み、それらの詳細内容を目視して最終的に必要なものを見つけるという 2 つのステップが必須である。データベース検索も大まかな絞り込みであり、その後には目視チェックが必須である。人による目視チェック不要の調査は極めて高度な人工知能が出現するまではあり得ない。

(1) IPC 記号による調査

20 世紀末までは、「人手による紙資料の調査は再現率 (精度) が高いが、コンピュータ検索 (データベース検索) は低い」と多くの人が言っていた。これは、人手による紙資料の調査では絞り込みの手段が IPC であったので目視件数がコンピュータ検索利用よりも多かったこと (および紙資料が明細書である調査では内容チェックが徹底できた一方で初期のデータベースでは明細書の参照が出来なかったこと) が上げられる。

(2) キーワードによる検索

技術内容指定の調査である限り、データベース検索を行っても最終的には目視チェックに必要な特許に絞り込むステップが必須である。検索だけで必要なものに絞り込むことは不可能である。

データベースでは検索キーとしてキーワードを使うため、IPC や FI 記号などよりも狭い範囲に絞り込むことができ便利である。しかし、狭く絞り込むほど必要なものが漏れる恐れが大きくなるので一定以上に絞り込んではいけない。

本題から少しそれるが、データベース検索ではノイズゼロ、漏れゼロはできないし、目標にしてはいけないと筆者は考えている。ノイズから教えられ、得た知識を使って別の条件で検索することで調査全体の再現率を上げることができる。逆に、検索で漏れゼロを目標にすると目視件数が多くなり過ぎて、コストがかかり過ぎる。

なお、データベース検索で IPC だけを使って検索するのであれば IPC 仕分け紙資料と同じ絞り込みとなる。漏れは減らせるだろうが、大量の目視チェックが必要になりコスト高になる。

(3) 概念検索

概念検索も大まかな絞り込みの手段であり、目視チェックが必須である。

難しそうな名前をしているために「概念検索はノイズや漏れの無い完璧な検索が行なえるべきである」とか、「概念検索は文法を理解すべきである」などの主張はせつかくの素晴らしいシステムの適用範囲を狭くすることである。機械に適した使い方をすれば良い。

概念検索の絞り込みの例：「半導体を冷却する」と「半導体で冷却する」の違いを人間は文法で理解できるが、コンピュータに理解させるのは困難である。現在の概念検索ではこの違いを取り扱うことはできない。

表 1 はこの 2 つの質問文で概念検索を行なった結果であるが、全く同じ結果である。これを見て、概念検索は使いものにならないと考える人がいるかもしれない。しかしこれは概念検索の特徴を生かしていない利用の結果と見るべきである。

表 2 は「半導体を冷却するフィン」で概念検索した結果であり、表 3 は「半導体のペルチェ効果で冷却する」の結果である。両者を見比べればそれぞれ検索目的 (=

表1 質問文「半導体で冷却する」と「半導体を冷却する」の概念検索結果

※システム：Sharesearch（シェアリサーチ）

項番	スコア	公報番号	発明の名称	出願人	IPC
1	1074	特開2001-203309	半導体モジュールの冷却構造	高周波熱錬株式会社	H01L23/473
2	1051	特開平06-268110	半導体装置	株式会社日立製作所	H01L23/34
3	1049	特開2000-180503	冷却装置	株式会社日立製作所	G01R31/26
4	1048	特開2012-256767	強制風冷式半導体冷却装置	東洋電機製造株式会社	H01L23/467
5	1046	特許02523771	電力変換装置	株式会社東芝	H02M1/00
6	1033	特開平08-037260	半導体冷却装置	株式会社日立製作所	H01L23/473
7	1031	特公平07-105422	半導体ウエハ載置台	東京エレクトロン株式...	H01L21/66
8	1029	特公平06-101524	半導体素子用冷却体	株式会社東芝	H01L23/473.800
9	1029	特開平06-310626	半導体チップ及び半導体集積回路...	株式会社日立製作所	H01L23/34
10	1027	特開平06-188345	半導体チップの冷却構造	三菱電機株式会社	H01L23/473.800

表2 質問文「半導体を冷却するフィン」の結果

項番	スコア	公報番号	発明の名称	出願人	IPC
1	1011	特開平05-048282	冷却フィン構造	株式会社三社電機製作所	H05K7/20
2	1008	特開平08-204369	冷却フィン構造	株式会社三社電機製作所	H05K7/20
3	980	特開平09-129794	半導体装置	株式会社日立製作所	H01L23/40
4	970	特開2005-348533	インバータ装置	富士電機システムズ株...	H02M7/48
5	970	特開2006-024661	冷却装置	株式会社日立製作所	H01L23/427
6	956	特許02962429	冷却装置	株式会社東芝	H01L23/427
7	955	特開2004-080656	電力変換装置	株式会社日立製作所	H02M7/48
8	932	特開平05-129487	半導体スタック装置	大阪電気株式会社	H01L23/40
9	951	特許02523771	電力変換装置	株式会社東芝	H02M1/00
10	944	特開平07-066334	ヒートシンク	日本電気株式会社	H01L23/36

表3 質問文「半導体のペルチェ効果で冷却する」の結果

項番	スコア	公報番号	発明の名称	出願人	IPC
1	847	特開2004-263325	ペルチェ素子を用いた冷暖房服	株式会社HI	A41D13/00
2	845	特開平08-274226	半導体装置	株式会社リコー	H01L23/38
3	840	特開平06-265640	ペルチェ冷却半導体検出器	セイコーインスツルメ...	G01T1/24
4	814	特開平06-294872	ペルチェ冷却半導体検出器	エスアイアイ・ナノテ...	G01T1/24
5	825	特開平08-070070	半導体装置	ヤマハ株式会社	H01L23/38
6	796	特開平09-139525	ペルチェ冷却ユニット構造	日立化成株式会社	H01L35/30
7	796	特開2004-165677	半導体装置	ヤマハ株式会社	H01L35/32
8	791	特開2001-223394	ペルチェ素子	有限会社東京医科電機...	H01L35/28
9	783	特開平10-223996	半導体レーザ装置	NECエレクトロニク...	H01S5/022
10	785	特開平11-214598	大規模集積回路(LSI)チップ...	青木武	H01L23/38

表4 質問文「半導体のペルチェ効果で冷却する衣服の結露対策」の結果

項番	スコア	公報番号	発明の名称	出願人	IPC
1	1109	特開2009-041142	偽装衣服	帝人ファイバー株式会社	A41D13/00
2	1104	特開2015-135932	ペルチェ冷却型ICパッケージ	アイアールスペース株...	H01L23/38
3	1094	特開2004-190884	ペルチェ式冷却装置及びその使用...	日東工業株式会社	F25B21/02
4	1094	特開2008-002131	非常用飲料水製造装置。	白石映慈	E03B3/28
5	1084	特開2003-176942	結露防止装置	三菱電機株式会社	F24F11/02
6	1050	特開平09-307030	結露防止機能付冷却装置	リコーエムックス株...	H01L23/34
7	1028	特開2004-293493	自己補給式水冷システム	株式会社日立製作所	F01P11/00
8	1007	特開平08-274519	空中線の結露防止装置	日本電気株式会社	H01Q1/02
9	1004	特開2012-018998	通信装置	富士通テレコムネット...	H05K7/20
10	1007	WO98/023235	保温・保冷チューブ及び保温・保...	三洋機工株式会社	A61F7/00.300

質問文)に近いものが出てきていることが分かる。

表2の中には「半導体で冷却する装置のフィン」も出てきているだろうが、それは目視チェックでノイズとして落とせば良いだけのことである。技術的には異なるが

表3も同様である。

なお、類似特許が多い場合には、さらに絞り込んだ質問文を使うのが良い。例えば、表4は「半導体のペルチェ効果で冷却する衣服の結露対策」で検索した結果である。



全てが目的に合ったものではないが、目的に近いものが散見される。

いかなる調査もデータベース検索も、まず最初に大まかな絞り込みを行い、最終的には目視チェックが必要なものだけを選出することが必要である。概念検索だけが特にノイズが多い訳ではない。ノイズや漏れの多寡はデータベースの選択と使い方による。

6 優れた製作者が優れた利用者ではない

ピアノ製造者が優れた演奏家ではない。車の設計者や製造者、販売者が優れたドライバーではない。利用者が使い方を研究して製造者にフィードバックすることで良いシステムに発展していく。数多く使って問題点を指摘し、改善を働きかけないとそのシステムは良いものにならない。その結果もっとも損するのは利用者である。

製造者もそれなりにがんばって良いものを作っているはずだが、利用者の視点での知識がないのだから一定以上は問題点に気付かないし改善しようがない。それを製造者の怠慢だと言っても良いものができるだけであり、被害者は利用者である。

ところが、ピアノ使用者の誰もが問題点や改善点に気付くわけではない。現在の利用法より上手な使い方を求めて研究的な利用を自ら行なった人だけがそれらに気付くことができる。問題点や改善点に気付いた人は製造者に知らせてやる必要がある。

例えば、概念検索には絞込検索という機能を持ったものが多い。これは IPC や F I 記号、さらにはキーワードや出願人などで限定した狭い範囲の特許を対象に概念検索を行なうものである。この機能は概念検索そのものではないので、システム開発者から見るとそれほど面白くないかもしれない。

しかし、利用者にとっては貴重なものである。なぜならば、一般の利用者は概念検索の細部には踏み込めないが、何とか他人よりも良い検索結果を出そうと努力するものである。それこそが検索者の価値を高めるものと言える。したがって、手を加える余地の大きいシステムが利用者、とくに専門調査員といわれる人にとって

は貴重なものである。絞込検索の併用は概念検索と検索式の組合せの意味を持つものであり、努力の注入できる余地のある部分である。このようなことは、概念検索を実際に調査ツールとして使っている利用者でないと分かり難い問題であろう。

僭越かもしれないが、概念検索システムの製造、開発者におかれては利用者からの良いシステムにするための提案を進んで吸収されることをお願いしたい。ローマは1日にしてならず。ピアノもチェンバロ（ハーブシコード）などの先立つ楽器があり、(たぶん) その利用者の声を聞き、改良を重ねてこそ現在の地位を勝ち得ているはずである。

7 進化する概念検索システム

いくつかの概念検索システムでは再現率向上のためにチューニングが行なわれており、より賢いシステムになってきている。

事例で紹介する。表5は4章でも紹介したものであるが、「インターネット競り取引で購入希望価格と最大許容値を入力できる」を質問文として2010年頃に検索した結果である。近い内容の特開平 10-078992 (以下 992 特許という) が2件目にある。表6は、質問文を「インターネット競り取引で購入希望価格と最大許容値を入力できる自動取引システム」というように「自動取引システム」を加えた結果である。992 特許が43番目に下がってしまっている。それほど重要とも思われない用語を不用意に追加しただけでこのように大きく変化している。当時はそんなものだと考えていたが最近大きく改善されている。

表7は2015年に「自動取引システム」を加えた質問文で概念検索した結果である。以前から近い内容の特許と考えていた 992 特許が、43番目付近ではなく最上位に来ている。チューニングの詳しい内容は分からないが、たいへん使い易いシステムになっていることが推察される。

TF・IDF は概念検索の基本であるが、再現率を向上させるために様々な工夫がなされて使い勝手のいいもの

表5 質問文「インターネット競り取引で購入希望価格と最大許容値を入力できる」の結果

※ 2013 年頃実施

項番	スコア	公報番号	発明の名称	IPC	出願人
1	320900	特開2002-056236	インターネット競売システム	G06F 17/60 316	松本 雅男
2	289467	特開平10-078992	自動競り方法	G06F 19/00	株式会社日立製作所
3	288632	特開2001-060237	自動競り方法	G06F 17/60 316	株式会社日立製作所
4	286396	特開2001-060238	競り方法およびその装置	G06F 17/60 316	株式会社日立製作所
5	278712	特開2001-319102	コンピュータ通信網を利用した販...	G06F 17/60 316	金 載吉
6	262493	特開2002-092387	インターネットを利用した競売及...	G06F 17/60 316	株式会社日本経営危機...
7	262491	特開平07-302287	競りシステム	G06F 19/00	株式会社富士通ゼネラル
8	260195	特開2002-183501	商品取引システム	G06F 17/60 314	堤 真人
9	259704	特開2001-306879	医薬品相互融通支援システム	G06F 17/60 316	株式会社ユニケソフト...
10	258893	特開2002-024580	商品取引システム、商品取引方法...	G06F 17/60 310	株式会社アルチェ

表6 質問文「インターネット競り取引で購入希望価格と最大許容値を入力できる自動取引システム」の結果

※ 2013 年頃実施；追加ワードで影響大

項番	スコア	公報番号	発明の名称	出願人
1	314263	特開2002-092438	商取引システム、販売システム、...	株式会社東芝
2	299409	特開2003-036370	商品取引方法及び商品取引システム	コンパステーパー株...
3	294537	特開2002-157480	電子商取引システム、電子商取引...	セピアコーポレーショ...
4	289243	特開2002-183501	商品取引システム	堤真人
5	287105	特開2002-024580	商品取引システム、商品取引方法...	株式会社アルチェ
6	283413	特開2001-319174	顧客補償電子商取引システムとそ...	福田幸雄
41	249753	特開2001-060238	競り方法およびその装置	株式会社日立製作所
42	254985	特開2001-060237	自動競り方法	株式会社日立製作所
43	256246	特開平10-078992	自動競り方法	株式会社日立製作所
44	255726	特開2002-007720	商品取引システム、方法及び記録...	ディーアイエスシステム...

表7 質問文「インターネット競り取引で購入希望価格と最大許容値を入力できる自動取引システム」の結果

※ 2015 年 6 月実施；追加ワードの影響が少ない

項番	スコア	公報番号	発明の名称	出願人
1	1615	特開平10-078992	自動競り方法	株式会社日立製作所
2	1615	特開2001-060237	自動競り方法	株式会社日立製作所
3	1597	特開2001-060238	競り方法およびその装置	株式会社日立製作所
4	1513	特開2006-252587	商品を配分するための処理を行う...	株式会社日立製作所
5	1407	特開2003-058773	インターネットを用いた競売シス...	小渡三佳矢
6	1378	特開2002-032617	自動競りシステム及び自動競り方法	株式会社日立製作所
7	1327	特開平07-302287	競りシステム	株式会社富士通ゼネラル
8	1312	特開2001-350955	取引仲介システム及び取引仲介方法	株式会社タマル
9	1280	特開2004-062693	商品取引システム、その方法及び...	日本電気株式会社
10	1278	特開2002-259742	物流情報システム	中塚治徳



に発展している。検索式はシステムが単純であり工夫の余地が少ない。これに対して概念検索は複雑な工程であるがゆえに工夫する余地も大きいといえる。

8 すでに存在する十分使えるシステム

特許調査に使える程度に十分な再現率を出せるシステムはすでに存在する。しかし全ての概念検索システムが同じように高い再現率を示すわけではない。検索式による検索ではシステムによる再現率の違いは少ないが、概念検索では非常に大きい。表8は、筆者が行なったすでに公表済みの実験²⁾の1つである。再現率が80%のシステムがある一方で、同じ質問文を使って0%のシステムがある。

また、使い方によって再現率に大きな違いが生じる。再現率を高めるような質問文の作成や使い方をしないと目的を達せられないことが多い。

したがって、概念検索と一口に評価するのではなく、区別して扱うべきであろう。ところがシステムによってそれほどの差のない検索式の経験を背景に概念検索をひと括りに見てしまうと訳が分からなくなってしまう。

これでは、せっかく良いシステムが存在するのに、利用者にとって不幸なことである。良いものは良いと正しく評価しよう。そうすれば、再現率の低いシステムも刺激を受けてやがて良いものになっていくことが期待でき、製造者も利用者も利益を得ることができよう。

9 終わりに

10年くらい前のことだが「最近のスマートホンってのは電話というよりパソコンだよ」と教えてくれた人がいた。当時私は携帯電話で満足しており、一部のスマホ利用者の使いこなしが難しいとか充電が面倒という話を聞き、知人の話しを理解しようとしなかった。最近使い始めたが、使ってみると確かにスマホはパソコン以上に便利な点がたくさんあり、手放せない道具になっている。

ところが、スマホは素晴らしいと教えてくれた知人は当時、「概念検索は何をやっているのか分からないしノイズや漏れが多くて使い物にならない。自分は使いたくない」と言っていた。この点は彼の認識違いだと考えている。

スマホのアプリには人がしゃべった内容を認知して応答・支援するものがある。最近はいろんな知的な人工頭脳の研究が進んでいるようだ。しかし、人の作る文章の文法解釈がずいぶんうまくいくようになった、という話はほとんど聞いたことがない。むしろ、ビッグデータの処理のように人とは異なる論理で動くものを耳にすることが多い。

本稿は、「なぜ私は概念検索の旗をふるのか？」という題名で書きだしたが、もう少し客観的な題名が良からうと考えて途中で変更したものである。

表8 生ごみ処理の脱臭関連特許の概念検索

※上位50件まで目視チェック、※ターゲットは11件

システム 検索対象	A		B	C	D	E	F
	要約	請求範囲	明細書	明細書	要約+請求範囲	明細書	明細書
抽出件数	4件	4	9	8	0	8	6
再現率	36%	36	82	72	0	72	55

参考文献

- 1) 概念検索の使いこなしに関する論文、六車正道、情報科学技術協会、情報の科学と技術 65 巻 3 号、141 ~ 147 (2015)
- 2) 「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」、P143、六車正道、発明推進協会、平成 25 年 5 月 (2013)
- 3) 概念検索はなぜ上手に検索できるのか?、六車正道、Japio YEAR BOOK 2008、2008.11 月