

## 検索式生成機能を有する特許公報検索システムの精度検証とその展望

柳堀恭子<sup>†</sup> 平塚三好<sup>†</sup> 谷川英和<sup>††</sup>

特許出願に必要な先行調査に、専門的な知識に頼らず機械的に調査できる Patent Search Assistant(PSA)を利用することができる。この PSA の課題として挙げられるのが、検索により示される特許コードの精度を上げることである。よって、この検索精度を検証し、検索結果に基づき、さらなる PSA の改良を促すものとする。

## Accuracy of Retrieval System and Its Prospects patent applications having the query generation

Kyoko Yanagihori<sup>†</sup> Mitsuyoshi Hiratsuka<sup>†</sup>  
Hidekazu Tanigawa<sup>††</sup>

Preceding the survey required a patent application, can be investigated without resorting to mechanical expertise Patent Search Assistant (PSA) can be used. This PSA is cited as a problem is to increase the accuracy of the code represented by a patent search. Thus, to verify the accuracy of this search, based on the results and encourage further improvement of the PSA.

### 1. はじめに

審査段階で拒絶される多くの特許は、同一または近似する先行特許文献が存在することにより拒絶されている。したがって、特許権取得の確率を上げるためには、特許出願する前段階での先行特許調査が不可欠であるといえる。

しかし、膨大な数にのぼる特許公開公報すべてに目を通し、完全に洩れのないように調査を行うことは不可能に近い。

さらに、特許検索式の作成については、調査者個人の経験やスキルに依存する。つまり、特許に用いられる用語の特殊性、および特許分類コード（以下、「特許コード」という）の理解の困難性から、熟練した特許調査者を除いて、適切な用語、および適切な特許コードを選択することは容易ではない。

そこで、適切な検索式を生成できる機能を有するシステムである Patent Search Assistant (PSA) が既に開発されている。なお、PSA は入力した用語の類義語、上位概念語、下位概念語を関連用語として提示する関連用語抽出機能と、入力した用語と選択した関連用語が出現する特許公開公報に付されている特許コードを提示する特許コード抽出機能を有する。

本論文では、この Patent Search Assistant (PSA) を用い、ユーザが任意に入力した検索キーワードから適切な特許コードを抽出することができるか、検索精度を検証し、その結果今後の特許調査支援につながる提案を行った。

### 2. 特許調査支援ツール(PSA)について

今回は Patent Search Assistant (PSA) を使用する。PSA は、特許検索時に使用する検索式の構築を支援するシステムであり、特許調査支援において唯一、特許検索式を自動生成するツールである。また、PSA は、以下の機能を有する。

---

<sup>†</sup> 東京理科大学  
Tokyo University of Science  
<sup>††</sup> IRD 国際特許事務所  
IRD Patent Office

### (1) キーワードを入力すると関連語を提案してくる

入力した各検索をキーとして、関連用語データベースを検索し、各検索キーワードに関連する用語が取得される(関連語提案機能)。なお、関連用語データベースは、約600万件の特許公開公報、およびWikipedia等のWebページから自動抽出した類義語の対、上位概念語と下位概念語の対からなるデータベースである。

そして、ユーザに提示された関連用語にチェックを入れることで、検索式に用いる用語を指定できる。



図1 関連語の提案

「データベース」「検索」というキーワードを入力した場合[a]

### (2) 特許コードを自動取得

検索用語からの関連用語を決定すると、PSAが勧める特許コードが自動的に取得される。PSAにおいて、「(入力された用語1+用語1の1以上の関連語)\*(入力された用語2+用語2の1以上の関連語)\*...\*(入力された用語n+用語nの1以上の関連語)」の用語のみの式を構成し、構成された式で検索される特許公開公報に付されている特許コードを抽出し、各特許コードに重要度を付し、重要度の順にソートして、特許コードを提示する。提示する特許コードは、IPC(国際特許分類)、Fターム、およびFIの3種類である。なお、Fタームは複数の観点(発明の目的、用途、材料など)から分類した日本国独自の特許分類コードであり、FIはIPCをさらに細分化した特許分類コードであり、日本国独自の特許分類コードである。

PSAにおいてコードの説明も表示されるため、必要となるコードを容易に取捨選択できる。

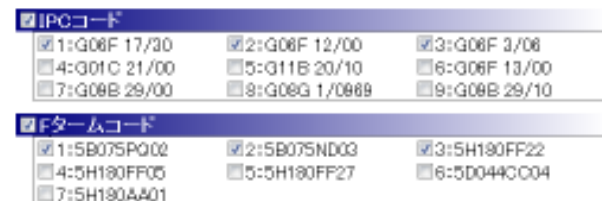


図2 特許分類コードの提案 [b]

### (3) 特許コードの組合せを表示



図3 検索式の構築 [c]

### (4) 特許公開公報が表示される

該当する特許公開公報の番号と名称が100件ずつ表示される。詳細な公報を調査する場合は、この特許公開公報の番号と名称から特定し、PSAと連携しているIPDLを利用して、さらに詳細な公報を見ることができる。

## 3. PSAによる特許コード抽出検証1

第一に、以下の手順で、PSAの特許コード抽出機能の評価を行った。

a), b), c) [http://www.ird-pat.com/corp/system\\_psa.html](http://www.ird-pat.com/corp/system_psa.html)

### 3.1 検証手順

特許庁が無償公開する特許電子図書館（以下、「IPDL」という）からキーワード入力  
 で得た特許公開公報を用意する。

同じキーワードを使い、PSAでの検索結果と比較する。

今回の検証ではIPCコードを利用した。

### 3.2 検証例

#### (1) 検証例 1

キーワードを「iPS細胞」とし、IPDLの特許データベースから公報を検索した。

該当公報は、35件であった。

一方、PSAで特許コード（IPCコード）を検索すると、PSAが特に推奨する特許コードとして、C12N 15/09、A61P 35/00、C12N 5/10の3つが抽出された。この上位3つのIPCコードの組合せで抽出された特許公開公報数は133件であった。この組合せでIPDL検索にて用意された35件のうちで合致するものは12件であった。

その他にPSAが推奨する特許コードとして上位3つ以外にA61P 43/00、A61K 38/00、C12Q 1/02、A61P 29/00、A61K 45/00、A61K 48/00、A61K 39/395が続いた。

#### (2) 検証例 2

キーワードを「LED電球」とし、IPDLの特許データベースから公報を検索した。

該当公報は、54件であった。

一方、PSAで特許コード（IPCコード）を検索すると、PSAが特に推奨する特許コードとして、F21Y 101/02、F21S 2/00の2つが抽出された。この上位2つのIPCコードの組合せで抽出された特許公開公報数は164件であった。この組合せでIPDL検索にて用意された54件のうちで合致するものは48件であった。

その他推奨する特許コードとして上位2つ以外にH01L 33/00、F21V 29/00、F21S 8/04、F21S 8/10、H05B 37/02、F21V 23/00、F21V 19/00、F21L 4/00が続いた。

#### (3) 検証例 3

キーワードを「耐震建築」とし、IPDLの特許データベースから公報を検索した。

該当公報は、40件であった。

一方、PSAで特許コード（IPCコード）を検索すると、PSAが推奨する特許コードとして、E04H 9/02、F16F 15/02、E04B 1/36、E02D 27/34、F16C 19/00の5つが抽出された。この5つのIPCコードの組合せで抽出された特許公開公報数は215件であった。この組合せでIPDL検索にて用意された40件のうちで合致するものは19件であった。

### 3.3 検証結果

色のついた部分は、特許公開公報にIPCコードが現れているにもかかわらず、PSAが推奨する上位10に示されたIPCコードにはなかったものを表している。

#### (1) 検証例 1 検索キーワード「iPS細胞」で検索した結果

公開公報 頻出順位	IPCコード	公開公報 頻出回数	PSAでの 優先順位
1	C12N 15/09	11	1
2	C12N 5/10	9	3
3	C12N 5/07	8	
4	A61L 27/00	6	
4	C12M 3/00	6	
6	C12M 1/00	5	
6	C12Q 1/02	5	6
8	A61K 48/00	3	9
8	A61P 43/00	3	4
8	C12N 5/06	3	
8	C12N 5/0735	3	
12	A61K 35/12	2	
12	A61K 35/28	2	
12	A61K 38/00	2	5
12	A61P 35/00	2	2

#### (2) 検証例 2 検索キーワード「LED電球」で検索した結果

公開公報 頻出順位	IPCコード	公開公報 頻出回数	PSAでの 優先順位
1	F21S 2/00	41	2
1	F21Y 101/02	41	1
3	H01L 33/00	20	3
4	F21V 29/00	18	4
5	F21S 8/04	9	5
5	F21V 19/00	9	9

5	F21V 23/00	9	8
8	F21S 8/10	8	6
9	F21V 3/04	6	
10	F21V 5/04	4	
10	H05B 37/02	4	7
12	F21V 7/04	3	
13	H01L 33/48	3	

(3) 検証例3 検索キーワード「耐震建築」で検索した結果

公開公報 頻出順位	IPC コード	公開公報 頻出回数	PSA での 優先順位
1	E04H 9/02	16	1
2	F16F 15/02	8	2
3	E04B 1/36	6	3
4	E04B 1/26	5	
4	E04B 2/56	5	
6	C22C 38/00	4	
6	E04B 1/58	4	
6	F16F 15/04	4	
9	C22C 38/06	3	
9	C22C 38/58	3	
9	E04B 1/24	3	
9	E04B 1/30	3	
9	E04B 1/34	3	
14	H04H 9/02	2	
14	E04B 1/18	2	
14	F16B 31/02	2	
14	F16B 5/02	2	
18	E02D 27/34	1	4
18	F16C 19/00	1	5

### 3.4 検証結果の考察

頻出する特許コード（IPC コード）の上位に、PSA 検索で抽出された特許コードが合致するかどうか焦点であった。

最上位頻出のコードについては、PSA でも抽出することができた。

PSA におけるコードの提案は、予めすべての公報から名詞句を抽出し、公報内における名詞句の出現頻度から、特定の公報における名詞句の重要度を算出しており、この重要度の高い公報に多く付されている IPC を順に表示するという手段がとられている。この重要度の高い公報に、ユーザが入力したキーワードが含まれていないと PSA が提案する上位 10 件の IPC コードから漏れる可能性がある。

また、検索式による公報検索時の公報を取得する方法では、形態素解析による全文検索機能を使って検索しているため、形態素の分割が、正しく行われない場合

（例えば  $iPS$  細胞 =  $iPS+$ 細胞ではなく、 $iPS$  細胞 =  $i+PS$  細胞と分けられる）は、正しい公報が取得されない。

## 4. PSA による特許コード抽出検証 2

第二に、以下の手順で、PSA の特許コード抽出機能の評価を行った。

### 4.1 検証手順

ランダムに特許公開公報を 100 件選ぶ。

各特許公開公報から頻出する上位 3 つの語句を検索のキーワードとする。

このキーワードを用いて PSA にて検索し、そして、さらに 100 件の特許公開公報の検証結果から、

- ①特許コードが抽出されず、公報も該当せず
- ②特許コードは抽出されなかったが、公報が該当した
- ③特許コードは抽出されたが、公報は該当せず
- ④特許コードが抽出され、公報が該当した

という 4 つの分類にわけた。

特許コードの太字は、特許公開公報にある特許コードと合致したものである。

#### 4.2 検証例

(1) 圧入位置決め方法 2005-001006 について

①特許コードが抽出されず、公報も該当せず にあたる例

公報の特許コード (IPC)	キーワード①	キーワード②	キーワード③
B23P 19/02	圧入位置	停止ポイント	圧入推力

①～③より抽出された IPC コード	PSA が出した公報数	ヒットの有無
—	0	無

①～③より抽出された IPC コード (全て)	PSA が出した公報数	ヒットの有無
—	0	無

(2) 切断装置およびそれに用いる回転刃ロールならびに刃物台ロール 2005-001029 について

②特許コードは抽出されなかったが、公報が該当した にあたる例

公報の特許コード (IPC)	キーワード①	キーワード②	キーワード③
B26D 1/38	回転刃ロール	刃物台ロール	切断

①～③より抽出された IPC コード	PSA が出した公報数	ヒットの有無
—	4	有

①～③より抽出された IPC コード (全て)	PSA が出した公報数	ヒットの有無
—	4	有

(3) コイル部品 2009-111111 について

③特許コードは抽出されたが、公報は該当せず にあたる例

公報の特許コード (IPC)	キーワード①	キーワード②	キーワード③
H01F 37/00 H01F 27/28	コイル	鏑部	コア

①～③より抽出された IPC コード	PSA が出した公報数	ヒットの有無
H01F 7/04 H01F 7/29	186	無

①～③より抽出された IPC コード (全て)	PSA が出した 公報数	ヒットの有無
H01F 17/04 H01F 27/29 H01F 30/00 <b>H01F 37/00</b> H01F 27/255 H02K 1/18 H01F 27/06 H01F 27/32 <b>H01F 27/28</b> H01F 27/24	330	無

- (4) 空気除菌装置 2007-222222 について  
 ④特許コードが抽出され、公報が該当した にあたる例 その 1

公報の特許コード (IPC)	キーワード①	キーワード②	キーワード③
A61L 9/16 F24F 6/00 F24F 6/04 A61L 9/01	空気除菌装置	気液接触部材	電解水

①～③より抽出された IPC コード	PSA が出した 公報数	ヒットの有無
<b>A61L 9/01</b> C02F 1/46	68	有

①～③より抽出された IPC コード (全て)	PSA が出した 公報数	ヒットの有無
<b>A61L 9/01</b> C02F 1/46 <b>A61L 9/16</b> <b>F24F 6/00</b> <b>F24F 6/04</b> A61L 9/015 A61L 9/00 B01D 46/00 A61L 9/14 B08B 3/04	73	有

- (5) 円筒ころ軸受および円筒ころ軸受用保持器 2008-232323 について  
 ④特許コードが抽出され、公報が該当した にあたる例 その 2

公報の特許コード (IPC)	キーワード①	キーワード②	キーワード③
F16C 33/46 F16C 19/26 F16C 33/66	円筒ころ	軸受	保持器

①～③より抽出された IPC コード	PSA が出した 公報数	ヒットの有無
<b>F16C 33/66</b> H02K 5/173 <b>F16C 19/26</b>	262	有

①～③より抽出された IPC コード (全て)	PSA が出した 公報数	ヒットの有無
<b>F16C 33/66</b> H02K 5/173 <b>F16C 19/26</b> F16C 33/58 F16C 33/80 F16C 33/76 F16C 41/00 <b>F16C 33/46</b> F16C 33/78 F16C 19/06	589	有

### 4.3 検証結果の考察

100 件の検索結果を①～④に分類したが、期待される成果は④の特許コードが抽出され、正しい検索式のもとに特許公開公報と合致することである。

検証例 (4) について

例に使用した特許公開公報に記載されていた特許コードは、A61L 9/16, F24F 6/00, A24F 6/04, A61L 9/01 の 4 件。3 つのキーワードから PSA が特に推奨する IPC コードは、A61L 9/01, C02F 1/46 の 2 件であった。特許公開公報記載の特許コードとの合致は A61L 9/01 の 1 件であったことから、再現率は  $1/2=0.5$ 、適合率は  $1/4=0.25$  であった。

PSA が推奨する IPC コードは、A61L 9/01, C02F 1/46, A61L 9/16, F24F 6/00, F24F 6/04, A61L 9/015, A61L 9/00, B01D 46/00, A61L 9/14, B08B 3/04 の 10 件であった。特許公開公報に記載の 3 件の特許コード全てを含み、再現率は  $4/10=0.4$ 、適合率は  $4/4=1.0$  であった。

検証例(5)について

例に使用した特許公開公報に記載されていた特許コードは、F16C 33/46, F16C 19/26, F16C 33/66 の 3 件。3 つのキーワードから PSA が特に推奨する IPC コードは、F16C 33/66, H02K 5/173, F16C 19/26 の 3 件であった。特許公開公報記載の特許コードとの合致は F16C 33/66, F16C 19/26 の 2 件であったことから、再現率は  $2/3=0.67$ 、適合率は  $2/3=0.67$  であった。

PSA が推奨する IPC コードは、F16C 33/66, H02K 5/173, F16C 19/26, F16C 33/58, F16C 33/80, F16C 33/76, F16C 41/00, F16C 33/46, F16C 33/78, F16C 19/06 の 10 件であった。特許公開公報に記載の 3 件の特許コード全てを含み、再現率は  $3/10=0.3$ 、適合率は  $3/3=1.0$  であった。

PSA が特に推奨する上位の特許コードで見た場合を第一段階、PSA が推奨する全ての特許コードで見た場合を第二段階とした。

そこで、100 件の結果をから、

第一段階では、①12 件、②22 件、③13 件、④53 件となった。

第二段階では、①12 件、②18 件、③10 件、④60 件となった。

このうち、①についてはキーワードの設定がおかしいなど、検索の方法に問題があると考えられるため省くこととし、

よって、②～④の条件で考えると、

④の特許コードが正しく抽出されて、該当の公報と合致する割合は、

第一段階では  $53/(100-12) \times 100 = 60.02\%$

第二段階では  $60/(100-12) \times 100 = 68.18\%$

多くの特許コードを候補と挙げる第二段階では、公報を見つけ出すのには 7 割近い正答を得たと考えられる。

問題は、該当公報が見つかったにもかかわらず、特定できる特許コードが抽出されないことがあること、公報に記載のある特許コードと同じコードが PSA で抽出されたにもかかわらず、該当公報が見つからなかったことである。

これは、PSA による特許コード抽出検証 1 における検証結果の考察と同様に、公報から抽出しておいた名詞句内に、今回入力したキーワードが含まれていない場合、②のように公報は該当したが、特許コードが抽出されないという現象が起きたといえる。

また、形態素分割が正しく行われえない場合には③のように、特許コードは抽出されたが、公報が該当しないという現象が起きたといえる。

## 5. まとめ・期待される改善点

PSA には、特許検索式を自動に生成してくれるという利点があり、そして、今回の検証の結果において、PSA は特許調査に必要な特許コードをかなり高い精度をもって提案することがわかった。

一方、PSA が特許調査支援だけにとどまらず、特許検索式から得られた特許公開公報情報にその特許が現在も権利が存続しているかなどの権利情報などが盛り込まれると、権利情報調査にも使えるという発展が望める。今後、特許出願から権利行使、他社からの防衛などの特許活動全体における特許調査活動を分析し、権利情報出力機能など、PSA に必要な機能の提案および実装を行っていききたい。

また、初心者用に十分な説明のある仕様書とともに研修会で使用するなどして、特許調査初心者にもさらに簡便な利用ができるようになることも期待したい。

## 6. おわりに

今回は、特許検索式を自動生成してくれることで、特許調査初心者でも使用できる PSA を使用したが、この PSA 以外にも、特許調査支援ツールが存在する。それらとの比較が、今後の課題でもある。

## 参考文献

- 1) 谷川英和, 河本欣司: 特許工学入門, 中央経済社(2003)
- 2) 高柳隆: 「特許調査」の基礎と応用, 工学図書(2006)
- 3) 谷川英和, 渡辺俊規, 増満光, 新森昭宏, 高木慎也, 難波英嗣: 特許記述言語 PML を用いた統合的特許構築システム, Japio 12/10 特許翻訳研究会第1回特許シンポジウム(2010).
- 4) [http://www.ird-pat.com/corp/system\\_psa.html](http://www.ird-pat.com/corp/system_psa.html)
- 5) 谷川英和, 渡辺俊規, 増満光, 難波英嗣: 特許調査から特許明細書作成を統合的に支援する特許構築システム, Japio YEAR BOOK, pp. 194-199 (2010)
- 6) 独立行政法人情報通信研究機構 知的財産(特許・商標)構築・活用のための情報通信基盤技術の研究開発 外部委託研究平成21年度報告書
- 7) 武藤晃: 特許調査における精度向上と検証方法, 特許・情報フェア&カンファレンス(2007)
- 8) 六道正道: 概念検索システムの現状と使いこなしの検討, 発明 2005.4,5月号, pp.1-14 (2005)
- 9) 酒井美里: 特許検索手法のマニュアル化と検索ノウハウの伝達, 情報管理 50(9), pp.560-577, (2007)
- 10) 難波英嗣, 竹澤寿幸, 乾孝司, 岩山真, 橋田浩一, 橋本泰一, 藤井敦: 特許検索履歴を用いたソースの自動構築, 言語処理学会第17回年次学会(2011)
- 11) 谷川英和, 渡辺俊規, 増満光, 難波英嗣: 特許調査の検索式提案システム, 日本知財学会年次学術研究発表会(CD-ROM), 8<sup>th</sup>:ROMBUNNO, 2G6(2010)