

ASEAN6知財庁データベースの検索性能検証

ASEAN データベースは検索に使えるか

Verification of ASEAN6 IP Office database



アジア特許情報研究会
中西 昌弘

2011年～ アジア特許情報研究会所属
2014年～ 日本特許データベースサービスセミナー講師
2018年～ 日本知的財産協会研修会講師
その他 JPO・JETRO・WIPO や民間会社主催セミナーでの講演

1 はじめに

筆者が所属するアジア特許情報研究会では日本貿易振興機構（JETRO）からの委託を受けて、2015年度より ASEAN 各国の知財庁データベースを調査し報告書を納入している。この報告書群（以降 JETRO 報告書⁽¹⁾と記す）では、各データベースの取り扱い説明だけではなく、どの程度の確かからしさで特許調査が可能なのかを、できる限り定量化した数字により考察した。

ひと口に特許調査と言っても様々な種類の特許調査が存在する。前記 JETRO 報告書では、

- ・ 狙いの技術範囲の特許群を抽出し
- ・ 権利範囲を特定する

ことをターゲットとした特許調査を前提としている。

仮に特許データベースを図1のような

- ・ A列には対象国で出願された案件の出願番号が記されており、
- ・ B列以降には各種書誌情報や発明の名称・要約・請求項・詳細な説明等々の「フィールド名」が配置され、
- ・ 縦横の交点セルに情報が入っている

スプレッドシートと想定したときに、特許情報収納庫としてのデータベースに求める必要条件を、

- ① 縦軸が長いこと
出願・発行された案件が網羅的に収録されていること
- ② 横軸が長いこと
特許調査に必要とするフィールドが網羅されていること

と

- ③ 交点セルに情報が入っていること

各案件の各フィールドに正しい情報が収録されていること
と考える。

	A	B	C	...	W	X	Y	Z
1	出願番号	出願日	公開番号		請求項	出願人	発明者	法律状態
2	AS1980-000001	※※※	※※※		※※※	※※※		※※※
3	AS1980-000002	※※※	※※※		※※※		※※※	※※※
...	...							
122345	AS1999-012344	※※※	※※※		※※※	※※※	※※※	※※※
122346	AS1999-012345						※※※	※※※
122347	AS2000-000001	※※※	※※※		※※※	※※※	※※※	※※※
122348	AS2000-000002		※※※		※※※	※※※		※※※
122349	AS2000-000003	※※※	※※※		※※※	※※※	※※※	※※※
122350	AS2000-000004	※※※	※※※			※※※	※※※	※※※
...	...							
987654	AS2015-045678	※※※	※※※			※※※	※※※	※※※
987655	AS2015-045679	※※※	※※※		※※※		※※※	

図1 仮想特許データベース

本稿では更に一步踏み込んで、特許情報検索システムとしてのデータベースの必要条件である、

- ④ 交点セルを検索できること

について、ASEAN6⁽²⁾の知財庁データベースを評価した結果を報告する。

2 知財庁データベースの検索性能

一般に電子データベースシステムでは、画面に表示されるレコード上のフィールドの内容を直接検索するのではなく、内部に保持する検索用インデクス内の情報を検索している。このためフィールドに収録された情報が正しくインデクスに「転記」されていない場合には、画面上で表示されるとおりの文字列情報を検索しても、その

レコードが検索されないという「検索異常」が起こりうる。

日本特許庁の FOPISER データベース⁽³⁾ におけるシンガポール (SG) 特許情報が、この現象の一例として挙げられる。FOPISER では画面に表示される情報はシンガポール知財庁 (IPOS) から提供された全文情報をもとに構成されているが、検索用インデクスに DOCDB (を管轄する欧州特許庁) から提供された書誌情報が収録されている。DOCDB の SG 案件の収録は非常に乏しく、せっかく IPOS から潤沢な情報が提供されても、この情報をほとんど検索できないという結果になっている。

筆者は ASEAN6 各国の知財庁データベースを使用し、

- ・発明の名称フィールドに含まれるテキスト情報
- ・要約フィールドに含まれるテキスト情報
- ・出願人名フィールドに含まれるテキスト情報
- ・IPC フィールドに含まれる IPC コード情報

のそれぞれが正しく検索されるかどうかを検証した。なお新興国知財庁が付与する IPC コードの正確性には今ひとつ不安が残るため、サブグループまで完備した IPC コードではなく、サブクラスまで、およびメイングループまでの IPC コードを検索したときの性能を評価した。

また知財庁データベースによる検索結果に、検索した文字列が含まれているかどうかだけでなく、検索でヒットしなかった「残りの案件群」には指定文字列が含まれていないことも検証することを狙いとしている。これを実現するために、指定調査期間の全案件について各案件の詳細画面で表示される文字列を抽出し、この情報を保存した検証用の DB (以降検証 DB と称す) を外部に作成し、庁 DB と検証 DB とでヒットしたレコードの出願番号が完全一致するかどうかを比較する方法により検索性能を検証した。

2.1 インドネシア

2.1.1 発明の名称・要約

庁 DB⁽⁴⁾ を使用して 2001 年～2010 年に申請された案件について、「Judul Invensi」フィールドから

各原語タームを検索した結果を表 1.1 に、「Abstrak」フィールドを検索した結果を表 1.2 に記す。

表 1.1 発明の名称検索結果検証

英語	原語	庁 DB	検証 DB	評価
bicycle	sepeda	414 件	同左	◎
camera	kamera	12 件	同左	◎
computer	komputer	88 件	同左	◎
engine	mesin	938 件	同左	◎
film	film	199 件	同左	◎
oil	minyak	631 件	同左	◎
radio	radio	187 件	同左	◎
television	televisi	33 件	同左	◎
water	air	872 件	同左	◎

表 1.2 要約検索結果検証

英語	原語	庁 DB	検証 DB	評価
bicycle	sepeda	333 件	同左	◎
camera	kamera	45 件	同左	◎
computer	komputer	270 件	同左	◎
engine	mesin	1,360 件	同左	◎
film	film	484 件	同左	◎
oil	minyak	1,108 件	同左	◎
radio	radio	331 件	同左	◎
television	televisi	33 件	同左	◎
water	air	2,912 件	2,911 件	○

詳細画面上では要約フィールドに「air」が表示されない P00200900731 が庁 DB による検索でヒットするという若干の異常が検出された。しかし画面表示された要約フィールドで使用されるいくつかの単語を検索してみると同案件がヒットすることから、まったく異なる案件の情報が検索インデクスに収録されているわけではなさそう。

2.1.2 出願人名

2001 年～2010 年出願の案件群を母集団として、ASEAN 地域への出願が多い 5 社について「Nama Pemegang Paten」フィールドの検索結果を検証した。

表 2 出願人名検索結果検証

出願人名	庁 DB	検証 DB	評価
PANASONIC	262 件	同左	◎
PFIZER	390 件	同左	◎
QUALCOMM	983 件	同左	◎
SAMSUNG	233 件	同左	◎
UNI-CHARM	418 件	同左	◎

同国の庁 DB では出願人名の検索も正しく動作しているようである。しかし例えば出願番号 W00200902786 の案件には、出願人「QUALCOM INCORPORATED」と社名末尾の M が欠落した文字列が収録されており、この案件は出願人「QUALCOMM」

を検索してもヒットしない。しかし本稿での報告は検索性能の検証であり、表記ゆれを考慮していないことに注意されたい。

2.1.3 IPC

2010年に出願された案件群を母集団として、同国知財庁で付与された件数の多い5種のIPCサブクラス、メイングループについて、詳細画面に表示される「IPC」フィールドの検索結果を検証した。

表 3.1 IPC サブクラス検索結果検証

IPC	庁 DB	検証 DB	評価
A61K	752 件	同左	◎
C07D	442 件	同左	◎
A61P	358 件	同左	◎
H04W	208 件	同左	◎
C07C	187 件	同左	◎

表 3.2 IPC メイングループ検索結果検証

IPC	庁 DB	検証 DB	評価
A61K31	440 件	同左	◎
A23L1	128 件	同左	◎
A61K9	98 件	同左	◎
A61P35	90 件	同左	◎
C12N15	90 件	同左	◎

同国の庁 DB では IPC 検索は正常に動作している。

2.1.4 検索性能検証まとめ

同国の庁 DB は Abstract 検索の僅かな差異以外は、全ての検索において正常に動作している。必要条件④を満たしていると考えられる。

2.2 マレーシア

2.2.1 発明の名称・要約

庁 DB⁽⁵⁾ を使用して 2001 年～2010 年に出願された案件について、「Title」フィールドから各タームを検索した結果を表 4.1 に、「Abstract」フィールドからの検索結果を表 4.2 に記す。

表 4.1 発明の名称検索結果検証

英語	庁 DB	検証 DB	評価
bicycle	13 件	同左	◎
camera	50 件	同左	◎
computer	298 件	同左	◎
engine	338 件	同左	◎
film	546 件	同左	◎
oil	725 件	同左	◎
radio	202 件	同左	◎
television	81 件	同左	◎
water	747 件	同左	◎

表 4.2 要約検索結果検証

英語	庁 DB	検証 DB	評価
bicycle	23 件	同左	◎
camera	206 件	同左	◎
computer	872 件	同左	◎
engine	745 件	同左	◎
film	1,654 件	同左	◎
oil	1,767 件	同左	◎
radio	579 件	同左	◎
television	169 件	同左	◎
water	3,979 件	同左	◎

発明の名称・要約ともに、すべてのタームが正しく検索された。

2.2.3 出願人名

続いて「Applicant/Name」フィールドの検索結果。

表 5 出願人名検索結果検証

出願人名	庁 DB	検証 DB	評価
PANASONIC	353 件	同左	◎
PFIZER	507 件	同左	◎
QUALCOMM	1,068 件	同左	◎
SAMSUNG	582 件	同左	◎
UNI-CHARM	485 件	同左	◎

「Applicant」についても、すべてのタームが正しく検索されることを確認した。

2.2.4 IPC

同国の庁 DB ではサブクラス (SC) やメイングループ (MG) までの IPC コードを検索する機能は用意されず、サブグループ (SG) まで全て完備した IPC コードしか検索できない。しかし検索する IPC コードとして任意の SC の下層の全コードを指定するための、言わば「アシスト画面」が用意されており、この機能を利用して SC 単位の検索検証を実施した。MG については検証を割愛した。

表 6 IPC サブクラス検索結果検証

IPC	庁 DB	検証 DB	評価
A61K	690 件	107 件	×
G06F	179 件	90 件	×
G01N	138 件	61 件	×
H01L	266 件	60 件	×
H04L	121 件	50 件	×

表のように、いずれの IPC についても庁 DB と検証 DB における検索結果が大きく乖離した。そこで庁 DB だけで検索された案件数、双方の DB で検索された案件数、検証 DB だけで検索された案件数の内訳を次表に記す。

表 7 IPC 検索結果内訳

IPC	庁 DB のみ	双方	検証 DB のみ
A61K	594 件	96 件	11 件
G06F	91 件	88 件	2 件
G01N	77 件	61 件	0 件
H01L	208 件	58 件	2 件
H04L	71 件	50 件	0 件

このように案件ごとの詳細画面に表示される IPC と、検索用インデクスに収録された IPC が全く異なるような状態。唯一の救いは詳細画面表示以上の数の案件が検索されることだけ。これではデータベースとしての必要条件④を満たしていないと言わざるを得ない。

2.2.4 検索性能検証まとめ

上記のように同国の庁 DB では、発明の名称・要約・出願人名のテキストフィールドは正しく検索されるが、IPC については表示される詳細画面上の IPC コードと検索用インデクスの IPC コードが一致していないことが判明した。このような状態では、IPC コードの機械検索により、査読すべき案件を絞り込むことは非常に危険であると言わざるを得ない。

2.3 フィリピン

2.3.1 発明の名称・要約

庁 DB⁽⁶⁾ を使用して 2001 年～ 2010 年に出願された案件について、「Title」フィールドから各タームを検索した結果を表 8.1 に、「Abstract」フィールドの検索結果を表 8.2 に記す。

表 8.1 発明の名称検索結果検証

英語	庁 DB	検証 DB	評価
bicycle	7 件	同左	◎
camera	12 件	同左	◎
computer	60 件	同左	◎
engine	177 件	同左	◎
film	151 件	同左	◎
oil	167 件	同左	◎
radio	74 件	同左	◎
television	21 件	13 件	○
water	361 件	同左	◎

表 8.2 要約検索結果検証

英語	庁 DB	検証 DB	評価
bicycle	8 件	同左	
camera	19 件	同左	◎
computer	99 件	同左	◎
engine	185 件	同左	◎
film	206 件	207 件	×
oil	267 件	同左	◎
radio	129 件	同左	◎
television	20 件	11 件	○
water	939 件	940 件	×

発明の名称・要約ともに、ターム「television」の件数に大きな乖離が見られる。庁 DB では文字列内の「TV」（略称）も「television」でヒットするようであり、乖離は全てこれが原因であった。

同国の庁 DB では発明の名称・要約検索時には単語単位の完全一致検索が実行される。スペースで区切られた文字列だけを「単語」と見なすのではなく、各種の記号も単語の区切り文字として扱われている。しかし同国の庁 DB では「'」は区切り文字としては動作しないようである。「film」の差の 1 件は PH/1/2013/501156 の要約文字列内の、「... 50% of the film's dry weight...」が「film」では検索されないことが原因

「water」の差の 1 件は、PH/1/2009/501138 の要約文字列内の「...in water...」が検索されないために発生したもの。

2.3.3 出願人名

続いて「Applicant」フィールド検索結果。

表 9 出願人名検索結果検証

出願人名	庁 DB	検証 DB	評価
PANASONIC	88 件	同左	◎
PFIZER	212 件	213 件	○
QUALCOMM	509 件	同左	◎
SAMSUNG	51 件	56 件	○
UNI-CHARM	139 件	同左	◎

PFIZER の差となった PH/1/2001/002394 の詳細画面では、

(71/73) Applicant

(E) ANDREW G. LEE : PFIZER GLOBAL RESEARCH AND DEVELOPMENT EASTERN POINT ROAD GROTON CONNECTICUT US

のように表示される。「Applicant」文字列全体のうち、

コロンより前の「ANDREW G. LEE」が「Applicant」フィールドに収録され、コロンの後ろの「PFIZER GLOBAL・・・」は「Applicant Address」フィールドの情報として扱われるために、「Applicant」フィールドの検索ではヒットしないようである。

SAMSUNG の差の 5 件についても、「SAMSUNG」が「Applicant Address」フィールドにしか収録されていないことが原因。

2.3.4 IPC

JETRO 報告書 PH 編にも記したように、同国の庁 DB では IPC サブクラス (SC) を検索する機能が用意されていない。しかし「A61K*」のようにワイルドカードを使用して前方一致検索することで、実質的に SC を検索することができる。2010 年に依頼された案件群を母集団として検証した結果を表 10 に記す。

この DB ではメイングループまでの IPC コードは検索できないため検証を割愛する。

表 10 IPC サブクラス検索結果検証

IPC	庁 DB	検証 DB	評価
A61K*	822 件	同左	◎
A61P*	521 件	同左	◎
C07D*	352 件	同左	◎
A23L*	139 件	同左	◎
C12N*	122 件	同左	◎

2.3.6 検索性能検証まとめ

上記のように同国の庁 DB では、発明の名称・要約・出願人名のテキストフィールドは僅かな区切り文字の違いを原因とする差異が確認されたが、ほぼ正確に検索され、必要条件④を満たしていると言える。しかし JETRO 報告書 PH 編に詳細を記したように、同国の DB では 2010 年以前に依頼された案件の要約収録率が低く、必要条件③に問題があることに注意されたい。

一般に新興国では「広めの IPC」を検索する機会が多い。メイングループまでの IPC を検索できないことは、必要条件④を満たしているとは言い難い。

2.4 シンガポール

2.4.1 発明の名称

庁 DB ⁽⁷⁾ を使用して 2001 年～2010 年に依頼された案件について、「Title」フィールドから各タームを検索した結果を表 11 に記す。なお同国の庁 DB の案件ごとの詳細画面には要約文字列が表示されないため、要約検索性能を比較検証することができない。このため同国については要約フィールドの検索性能検証は割愛する。

また同国の庁 DB でのテキスト検索時には、他国のように単語単位の完全一致検索は実行できないため、「Contains」オペレータを使用した部分一致検索を実行したものである。

表 11 発明の名称検索結果検証

英語	庁 DB	検証 DB	評価
bicycle	7 件	同左	◎
camera	81 件	同左	◎
computer	579 件	同左	◎
engine	503 件	同左	◎
film	1,286 件	同左	◎
oil	1,156 件	同左	◎
radio	503 件	同左	◎
television	49 件	同左	◎
water	1,147 件	同左	◎

表のように「Title」フィールドの検索は正しく動作していることが確認できた。

2.4.2 出願人名

続いて「Applicant/Proprietor」フィールド検索結果。

表 12 出願人名検索結果検証

出願人名	庁 DB	検証 DB	評価
PANASONIC	285 件	同左	◎
PFIZER	550 件	同左	◎
QUALCOMM	2,295 件	同左	◎
SAMSUNG	445 件	同左	◎
UNI-CHARM	260 件	同左	◎

2.4.3 IPC

2010 年に依頼された案件群を母集団として IPC サブクラス (SC) を検索した結果を表 13.1 に、メイングループ (MG) の検索結果を表 13.2 に記す。同国の庁 DB では「IPC No.」の検索時に「Starting with」オペレータを使用すると、複数の IPC が列記された場合の 2 個目以降の IPC が検索されないため、文字列検

索時と同様に「Contains」オペレータを使用した。また JETRO 報告書 SG 編に詳細を記したように同国の庁 DB では約 93% の IPC コードは SC と MG 間に 1 個の半角スペースが含まれている。このため表に記したように MG 検索においては半角スペースを含んだ文字列を検索した。

表 13.1 IPC サブクラス検索結果検証

IPC	庁 DB	検証 DB	評価
A61K	1,499 件	同左	◎
A61P	836 件	同左	◎
C07D	630 件	同左	◎
G06F	316 件	同左	◎
H01L	622 件	同左	◎

表 13.2 IPC メイングループ検索結果検証

IPC	庁 DB	検証 DB	評価
A61K△31	925 件	926 件	△
H01L△21	458 件	同左	◎
A61P△35	230 件	同左	◎
A61K△39	200 件	同左	◎
A61K△9	188 件	同左	◎

MG 検索で僅かな問題が発生している。出願番号 2011086022 の案件の詳細画面上では「A61K△31/47」が表示されるにも関わらず、この案件は「A61K△△31/47」のように SC・MG 間に 2 個の半角スペースを挟まないとヒットしない。検索用インデクスには「A61K△△31/47」が入っている様子。

2.4.4 検索性能検証まとめ

発明の名称・出願人名の文字列検索、IPC 検索ともに、ほぼ問題なく動作していることがわかった。必要条件④は満たしているものの、JETRO 報告書 SG 編に記したように、庁 DB には IPC が付与されていないレコードが大量に収録されている。そもそも必要条件の③に問題があることに注意されたい。

2.5 タイ

2.5.1 発明の名称・要約

庁 DB⁽⁸⁾ を使用して 2001 年～2010 年に出願された案件について、「Title」フィールドから各タームを検索した結果を表 14.1 に、「Abstract」フィールドの検索結果を表 14.2 に記す。タイ語文字列は必ずしも分かち書きされていないため、検索には「Comparison」オペレータに「like」を設定することで部分一致モードでの検索を実施した。Google 翻訳を使用して water

をタイ語に翻訳すると「น้ำ」と 1 文字のタームに変換される。この「น้ำ」を検索すると桁外れの件数がヒットする。この集合は必ずしも全てが water について記されたものではなく、たまたま同じ表音文字を使用した、別の内容の案件であると推測している。しかし本稿は庁 DB の「文字列検索」の性能を検証するものであり、「内容検索」検証ではないことをご容赦願いたい。

表 14.1 発明の名称検索結果検証

英語	原語	庁 DB	検証 DB	評価
bicycle	รถจักรยาน	259 件	同左	◎
camera	กล้อง	48 件	同左	◎
computer	คอมพิวเตอร์	183 件	同左	◎
engine	เครื่องยนต์	754 件	同左	◎
film	ฟิล์ม	581 件	同左	◎
oil	น้ำมัน	661 件	同左	◎
radio	วิทยุ	98 件	同左	◎
television	โทรทัศน์	57 件	同左	◎
water	น้ำ	2,587 件	同左	◎

表 14.2 要約検索結果検証

英語	原語	庁 DB	検証 DB	評価
bicycle	รถจักรยาน	264 件	同左	◎
camera	กล้อง	173 件	同左	◎
computer	คอมพิวเตอร์	541 件	536 件	△
engine	เครื่องยนต์	1,429 件	1,430 件	△
film	ฟิล์ม	1,459 件	1,451 件	△
oil	น้ำมัน	1,670 件	1,659 件	△
radio	วิทยุ	216 件	217 件	△
television	โทรทัศน์	108 件	109 件	△
water	น้ำ	9,825 件	9,757 件	△

Title については庁 DB と検証 DB の検索結果が完全一致したが、Abstract については bicycle と camera 以外は検索された集合が異なるという結果。computer 以降のタームについて、庁 DB だけで検索された案件数、双方の DB で検索された案件数、検証 DB だけで検索された案件数の内訳を表 15 に記す。

表 15 要約検索結果内訳

英語	原語	庁 DB のみ	双方	検証 DB のみ
computer	คอมพิวเตอร์	5 件	536 件	0 件
engine	เครื่องยนต์	2 件	1,427 件	3 件
film	ฟิล์ม	8 件	1,451 件	0 件
oil	น้ำมัน	13 件	1,657 件	2 件
radio	วิทยุ	1 件	215 件	2 件
television	โทรทัศน์	0 件	108 件	1 件
water	น้ำ	77 件	9,748 件	9 件

「庁 DB のみ」に分類された案件は、いずれも詳細画

面に要約文字列が表示されないために、検証 DB では各タームが検索されなかった案件。逆に「検証 DB のみ」に計数された案件群は詳細画面上の要約文字列内に、所定の原語タームが存在するものの、庁 DB の検索ではヒットされなかった案件。

文字列検索性能に若干の懸念は残るものの、庁 DB で検索できなかった案件数は全体の 0.11% と非常に小さく、庁 DB の文字列機械検索性能を一概に否定するものではないと考える。

2.5.2 出願人名

続いて「Applicant」フィールドの検索結果。

表 16 出願人名検索結果検証

英語出願人名	原語出願人名	庁 DB	検証 DB	評価
PANASONIC	พานาโซนิค	436 件	同左	◎
PFIZER	ไฟเซอร์	448 件	同左	◎
QUALCOMM	ควอลคอมม์	311 件	同左	◎
SAMSUNG	ซัมซุง	127 件	同左	◎
UNI-CHARM	ยูนิ-ชาร์ม	555 件	同左	◎

表に記したように出願人文字列は正常に検索されることが確認できた。

タイ文字は表音文字であり、同国の庁 DB 上では国外出願人名は、耳で聞こえた音をそのまま表音文字で記している。たとえば日本特許において出願人名「フランス国立科学技術センター」を「サントゥルナショナルドゥルシエルシュシアンティフィック」と表記しているのと同様であり、国外出願人表記には夥しい揺れが含まれている。本稿は文字列検索性能の検証報告。表に記した「原語出願人名」は、多数の表記の中の一例であって。その出願人名を網羅的に表していないことに注意されたい。

2.5.3 IPC

2010 年に出願された案件群を母集団として IPC サブクラスを検索した結果を表 17.1 に、メイングループの検索結果を表 17.2 に記す。

表 17.1 IPC サブクラス検索結果検証

IPC	庁 DB	検証 DB	評価
A61K	927 件	同左	◎
C07D	397 件	同左	◎
A61P	198 件	同左	◎
C07C	183 件	同左	◎
B65D	171 件	同左	◎

表 17.2 IPC メイングループ検索結果検証

IPC	庁 DB	検証 DB	評価
A61K△31	506 件	同左	◎
A61K△8	183 件	同左	◎
A23L△1	112 件	同左	◎
A61F△13	100 件	同左	◎
A61K△9	98 件	同左	◎

上表のように IPC 文字列の検索は正常に動作している。

2.5.4 検索性能検証まとめ

発明の名称・要約の文字列検索結果が完全一致したわけではないが、詳細画面に表示されるものの検索できない案件は 0.1% 程度と非常に僅かである。出願人検索については検証した限りでは完全一致していた。IPC の検索にも問題がなく、同国の庁 DB は必要条件④を満たしていると考ええる。

2.6 ベトナム

同国知財庁では公開された特許・実案をすべて収録した IPLib⁽⁹⁾ と、登録に至った案件だけを収録する DigiPat の 2 種類のデータベースを運営している。ここでは IPLib の検証結果を報告する。

2.6.1 発明の名称・要約

IPLib では「Title of invention」フィールド選択肢を使用して発明の名称を検索することはできるが、要約だけを対象とした検索はサポートされていない。代わりに「Title or Abstract Text」選択肢が用意され、Title と Abstract のいずれかに収録された文字列を検索することができる。

ベトナム語は「単語」の文字数が短く、仮に英語の場合にひとつの単語で表現するものを複数の単語（連語・フレーズ）で表すことが多い。IPLib では連語の「má y ành」をそのとおりに検索窓に入力して検索すると、「má y」と「ành」のいずれかを含んだ文字列を検索（OR 検索）する。このため表に記したように、連語を検索するときにはダブルコーテーション記号で挟んだ文字列を検索した。

表 18 に IPLib を使用して 2001 年～2010 年に 出願された案件について、「Title」フィールドから各タームを検索した結果を記す。

表 18 発明の名称検索結果検証

英語	原語	庁 DB	検証 DB	評価
bicycle	"xe đạp"	72 件	同左	◎
camera	"máy ảnh"	3 件	同左	◎
computer	"máy vi tính"	2 件	同左	◎
engine	"động cơ"	527 件	同左	◎
film	"phim ảnh"	1 件	同左	◎
oil	dầu	280 件	同左	◎
radio	radio	3 件	同左	◎
television	"vô tuyến"	141 件	同左	◎
water	nước	866 件	同左	◎

IPLib では「Abstract」フィールドの文字列を正しく検索することができない。次の表 19 は各タームを「Title of invention」で検索した結果件数を TI 列に、「Title or Abstract Text」での結果を「TI or AB」列に記したものの。

表 19 名称・要約検索結果比較

英語	原語	TI	TI or AB
bicycle	"xe đạp"	71 件	71 件
camera	"máy ảnh"	3 件	3 件
computer	"máy vi tính"	2 件	2 件
engine	"động cơ"	527 件	527 件
film	"phim ảnh"	1 件	1 件
oil	dầu	280 件	280 件
radio	radio	3 件	27 件
television	"vô tuyến"	141 件	141 件
water	nước	865 件	865 件

このようにターム「radio」以外は、検索するフィールドを Abstract まで広げても検索結果件数が増加していない。何かの異変が起きていると考えられる。そこで各種のベトナム語タームを、「Title of invention」および「Title or Abstract Text」で検索し、双方の結果を比較してみた。結果を表 20 に記す。

表 20 各種 VN ターム検索結果比較

日本語	原語	TI	TI or AB
スクーター	"xe tay ga"	1 件	5 件
オートバイ	"xe máy"	167 件	167 件
トラック	"xe tải"	5 件	5 件
アクセル	"ga"	37 件	86 件
ブレーキ	"phanh"	80 件	134 件
ビン	"chai"	38 件	77 件
カン	"lon"	12 件	17 件
スプーン	"muỗng"	1 件	1 件
ナイフ	"con dao"	0 件	1 件
フライパン	"chảo"	6 件	6 件
冷蔵庫	"tủ lạnh"	25 件	25 件
電子レンジ	"lò vi sóng"	1 件	1 件
耳	"tai"	15 件	96 件
鼻	"mũi"	39 件	39 件
口	"miệng"	213 件	213 件
歯	"răng"	252 件	252 件
肩	"vai"	0 件	149 件
心臓	"tim"	36 件	200 件

薄緑に色付けした部分のようにアルファベット 26 文字だけで構成された単語・連語は、検索対象フィールドを Abstract まで広げることによって検索結果件数が増加している。一方声調記号等の記号が付与された文字を含む単語・連語は、TI 列と TI or AB 列で検索結果件数が増加していないことがわかる。

このように記号付きのベトナム文字を含んだタームは、「Abstract」フィールドからは検索できない。IPLib はデータベースの必要条件④を満たしていないと判断する。

2.6.2 出願人名

IPLib のレコードは出願番号単位で構成され、1 個のレコード内に公開情報と登録情報が混在している。案件ごとの詳細画面上では、公開された段階では「(71) Applicant Name」フィールドに出願人名が記されているが、案件が登録されるとこのフィールドが表示されなくなり、代わりに「(73) Right Holder Name」フィールドが表示される。しかし、これは表示を消しているだけであり、「(71) Applicant Name」情報がレコード上から削除されるわけではない。登録特許についても「Applicant Name」フィールドの検索は可能である。しかし画面上に表示されない限り、検索結果が正しいかどうかを比較検証することができない。

このため VN 国知財庁 DB については、出願人名検索性能の検証は割愛する。

2.6.3 IPC

2010 年に出願された案件群を母集団として IPC サブクラス (SC) を検索した結果を表 21.1 に、メイングループ (MG) の検索結果を表 21.2 に記す。JETRO 報告書 VN 編にも記したように、庁 DB 上の IPC 表記は A01B△2/34 のように SC と MG の間に半角スペースが挟まれた文字列が最も多いことがわっている。そこで MG 検索においては SC・MG 間に半角スペースを挟んだ文字列を検索する。さらに連語(フレーズ) 検索をさせるために "A01B△2/34" とダブルコーテーションで囲った文字列を検索する。

表 21.1 IPC サブクラス検索結果検証

IPC	庁 DB	検証 DB	評価
A61K	612 件	同左	◎
A61P	421 件	同左	◎
C07D	397 件	396 件	○
H04W	234 件	同左	◎
H04L	143 件	同左	◎

表 21.2 IPC メイングループ検索結果検証

IPC	庁 DB	検証 DB	評価
"A61K△31"	452 件	同左	◎
"A61P△35"	130 件	同左	◎
"C07D△401"	113 件	同左	◎
"A01N△43"	98 件	97 件	○
"A61P△25"	96 件	同左	◎

上表のように SC、MG 検索の双方とも、若干の差異が発生している。これはいずれも出願番号 1-2010-03408 の案件についての差異。この案件の詳細画面を庁 DB で表示させても、書誌情報が全て空白であり、検証 DB では当該案件の情報を抽出できないことが要因。それ以外は庁 DB の検索性能に問題はない。

2.6.4 検索性能検証まとめ

前記のように同国の庁 DB では検索するタームに、声調記号等を含んだ文字が含まれると Abstract フィールドを検索しても何もヒットしないという重大な問題が潜んでいる。調音文字を含まないベトナム単語は稀であり、ほとんどの単語を検索できないと言っても言い過ぎではないほど。要約を検索できないデータベースでは、査読すべき案件を機械検索するにあたり大きなダメージと言わざるを得ない。

3 ASEAN データベースは検索に使えるか

最初に記したように、我々ユーザが知財庁データベースに期待する役割には、単なる特許情報の収納庫としてだけでなく、収納（収録）された情報を正しく検索できることも含まれている。いくら情報を収録していても、その情報を正しく検索できないようでは電子データベースと呼ぶことはできない。この観点から実施した検証結果を一覧にまとめると次のようになる。

表 22 各国データベース必要条件④一覧

国名	名称	要約	出願人名	IPC SC	IPC MG
ID	インドネシア	○	○	○	○
MY	マレーシア	○	○	○	×
PH	フィリピン	○	○	○	×
SG	シンガポール	○	—	○	○
TH	タイ	○	○	○	○
VN	ベトナム	○	×	—	○

国ごとの項で説明したように、シンガポールの要約や、ベトナムの出願人名は検索結果を比較検証することができなかったため、表では判定を「—」とした。各国それぞれ僅かな問題は発見されたものの、検索システムとしての必要条件④を満たしていないと結論づけられるのは、マレーシアの IPC 検索、フィリピンの IPC メイングループ検索、ベトナムの要約文字列検索であった。

4 ASEAN データベースの今後

国によっては毎年のようにデータベースがリニューアルされ、「調査条件」が徐々に進展している。

検索する日によって結果が異なったインドネシアも 2018 年 4 月のリニューアル以降は、検索結果に対する信頼性が大きく向上した。マレーシアは 2018 年 12 月のリニューアルにより検索エンジンが正常化した。IPC 検索についてはまだまだ不安・不満が残る状態。2018 年 10 月にリニューアルしたフィリピンサイトは WIPO のベースエンジンを同国用にローカライズしたもの。まさか WIPO エンジンで IPC メイングループ検索ができないとは思えず色々調べているが、未だに検索する方法が見つからないところ。

10 年ほど進化がなかったベトナムの IPLib。現在ベトナム知的財産研究所 (VIPRI) ⁽¹⁰⁾ により、新しい検索エンジンの開発中である。現状のプロトタイプを試してみたところ、声調記号が付与された「ベトナム文字」であっても要約を正常に検索できることを確認済みである。

ASEAN 各庁データベースの今後の進化に期待したい。

補足・参考文献

(1) JETRO 報告書

ID 編 : https://www.jetro.go.jp/ext_images/world/asia/idn/ip/pdf/search_ip_communique2018.pdf

MY 編 : https://www.jetro.go.jp/ext_images/world/asia/my/ip/pdf/search_ip_communique2017.pdf

PH 編 : https://www.jetro.go.jp/ext_images/world/asia/ph/ip/pdf/search_ip_communique2018.pdf

SG 編 : https://www.jetro.go.jp/ext_images/world/asia/sg/ip/pdf/search_ip_communique2017.pdf

TH 編 : https://www.jetro.go.jp/ext_images/world/asia/th/ip/pdf/search_ip_communique2017.pdf

VN 編 : https://www.jetro.go.jp/ext_images/world/asia/vn/ip/pdf/search_ip_communique2017.pdf

(2) ASEAN6

ASEAN10 か国の中である程度の規模で特許情報が収録されたデータベースを運営するインドネシア・マレーシア・フィリピン・シンガポール・タイ・ベトナムの6か国を表す。

(3) FOPISER データベース

<https://www.foreignsearch.jpo.go.jp>

(4) インドネシア知財庁 DB

<https://pdki-indonesia.dgip.go.id/>

(5) マレーシア知財庁 DB

<https://iponline2u.myipo.gov.my/myipo/www/>

(6) フィリピン知財庁 DB

<http://onlineservices.ipophil.gov.ph/wopublish-search/public/patents>

(7) シンガポール知財庁 DB

<https://www.ip2.sg/RPS/WP/CM/SearchAdvanceP.aspx>

(8) タイ知財庁 DB

<https://patentsearch.ipthailand.go.th/DIP2013/complexsearch.php?lang=en>

(9) ベトナム知財庁 DB IPLib

<http://iplib.noip.gov.vn/WebUI/WSearchPAT.php>

(10) ベトナム知的財産研究所

<http://english.vipri.gov.vn/>