

ASEAN特許文献のテクニカルターム検証

タイ語、ベトナム語、インドネシア語のテクニカルターム収集による検証

○佐野浩太郎¹⁾, 中西 昌弘²⁾

株式会社 IHI¹⁾, オリンパス株式会社²⁾

〒135-8710 東京都江東区豊洲三丁目1番1号 豊洲 IHIビル

Tel: 03-6204-7093 FAX: 03-6204-8651

E-mail: koutarou_sano@ihi.co.jp

Study of Technical Term Described in ASEAN Countries Patent Gazette: Study by Collecting Technical Term of Thai, Vietnamese and Indonesian

SANO Kotaro¹⁾, NAKANISHI Masahiro²⁾

IHI Corp.¹⁾, Olympus Corp.²⁾

Toyosu IHI Building, 1-1, Toyosu 3-chome, Koto-ku, Tokyo 135-8710 Japan

Tel: +81-3-6204-7093 FAX: +81-3-6204-8651

E-Mail: koutarou_sano@ihi.co.jp

【発表概要】

従来、英語を公用語としない新興国の特許調査を行う場合、商用データベースで英語キーワード、あるいは、IPC (International Patent Classification) を用いて検索されてきた。しかし、上記手法では他国にファミリー出願がされていない新興国特許の検索が困難である。そのため、侵害予防調査などを行う場合、原語キーワードを用いた検索で補完する必要がある。

本報告は、原語キーワードを用いた検索の網羅性を向上させる提案に関する。具体的には、ファミリー型(ファミリーを1レコードとして構成された)商用データベースの検索結果から得られたファミリー出願(タイ、ベトナム、インドネシア)から原語キーワードを抽出する。そして、抽出した原語キーワードからの異表記、あるいは、ワイルドカードを用いた表記の検証例を紹介する。さらに、IPCのみを用いた検索で生じる検索漏れの現状を示す。

【キーワード】

新興国, 特許調査, テクニカルターム, キーワード検索, 異表記, ワイルドカード, タイ語, ベトナム語, インドネシア語

1. はじめに

近年、ASEAN 諸国は、日本企業の投資先として注目されている。例えば、2011年の日本対ASEAN直接投資額では中国向けの投資額を上回っている[1]。特許出願の件数については、タイ、ベトナム、マレーシア、インドネシアにおいて、日本からの出願件数が大きく伸びている。特に、2013年のタイ、ベトナムにおける海外からの出願件数に関しては、日本が最も多い。また、ASEANの全ての国における海外からの出願件数では、米国、日本、ドイツが、上位に入っている傾向がみられる[2]。このような背景などから、侵害防止調査の重要度・必要性が認識されてきており、ASEAN 諸国における特許調査の必要性が高まっている。

従来、英語を公用語としない新興国の特許調査を行う場合には、ファミリー型（ファミリーを1レコードとして構成された）商用データベースで英語キーワード、あるいは、IPC (International Patent Classification) を用いた検索が行われてきた。しかし、英語キーワードを用いた検索に関しては、他国にファミリー出願がされていない新興国特許の検索が困難である。また、IPC を用いた検索では、新興国の特許公報へのIPC付与率が十分ではない（付与なし又は誤付与がある）ことから、他国にファミリー出願がされていない新興国特許のIPCによる検索が困難である。そのため、侵害予防調査などの検索漏れがない調査を行う場合には、原語キーワードを用いた調査で補完する必要がある。

また、原語キーワードに関しては、機械翻訳から抽出する手法、IPCを用いた検索結果から抽出する手法、ファミリー型商用データベースの検索結果から得られたファミリー出願から抽出する方法が紹介されている[3]。

2. 目的

本報告は、タイ、ベトナム、インドネシアの特許文献の調査において、原語キーワードを用いた検索の網羅性を向上させる提案に関する。具体的には、抽出した原語キーワードから、機械翻訳により、単語の最小構成単位を検討した異表記の検証例を紹介する。加えて、ワイルドカードを用いた表記の検証例を紹介する。さらに、IPCのみを用いた検索で生じる検索漏れの現状を示す。

3. 原語キーワードを用いた検索式検討

タイ、ベトナム、インドネシアの特許庁データベースを用いた検索式の検討方法及び検討結果を以下に示す。

3-1 検証方法

以下の手順で原語キーワードを用いた検索式を検討した。

(1)対象国を収録し、且つファミリーを検索できる特許データベースを使用し、国コードを対象国に指定し、IPCによる検索を行った（使用データベース：PatBase®）。本報告では、単語及び連語の双方について検証可能な点から、ターボチャージャを検証事例とした。

(2)ヒット公報のファミリーに記載された発明の名称から原語キーワードを抽出した。上記が収録されていない場合には、公報番号を抽出し、各国特許庁データベースで番号検索を行い、ヒット公報から原語キーワードを抽出した。発明の名称から原語キーワードを抽出する際、使用ツール：Google 翻訳）を用いた。なお、上記手順に関しては参考文献で紹介されている[3]。

(3)上記(2)で抽出した原語キーワードについて、さらに機械翻訳を行い、単語の最小構成単位を検討した(図1参照)。加えて、タイ語、ベトナム語に関しては、抽出した原語キーワードに関して、声調

記号を除いたものを異表記として収集した。

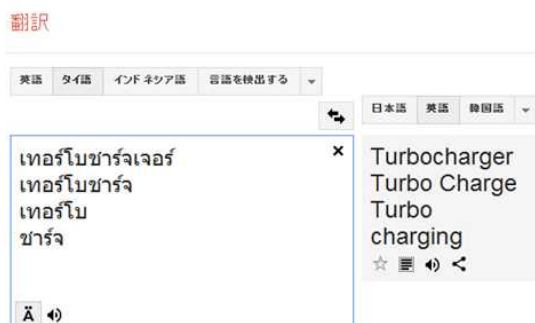


図 1. 単語の最小構成単位の検証

(4)上記(2)及び(3)で収集した原語キーワードを用いて、各国特許庁データベースで検索し、ヒット件数を確認した。さらに、ワイルドカード記号を用いた表記について検索を行い、網羅性を向上させる検索式の検討を行った。

3-2 タイ語の検証

ターボチャージャを対象とした検証結果を表 1 に示す。声調記号の有無に関わらず同じ訳があてられた用語を検索すると、それぞれのヒット件数に差がみられた。例えば、เทอร์โบชาร์จ と เทอร์โบชาร์จの場合、訳はどちらも Turbo Charge であるが、ヒット件数が、前者が 14 件に対して後者が 1 件と相違した。

また、連語の場合、それぞれの用語の順序を逆にしても、同一の訳があてられる場合があった。この場合、それぞれの用語を検索すると、ヒット件数に差がみられた。例えば、เทอร์โบ ชาร์จ และ ชาร์จ เทอร์โบの場合、訳はどちらも Turbo Charge であるが、ヒット件数が、前者が 13 件に対して後者が 0 件と相違した。

表 1. タイ語異表記の検証

タイ語	英語(訳語)	DIPHTH ヒット 件数
เทอร์โบชาร์จ	Turbo Charger	1 11
เทอร์โบชาร์จเจอร์	Turbocharger	
เทอร์โบชาร์จ	Turbo Charge	13
เทอร์โบชาร์จ		1
เทอร์โบชาร์		14
เทอร์โบชาร์		1
เทอร์โบชาร์		1
ประจุแบบเทอร์โบ		0
ประจุเทอร์โบ		0
ชาร์เทอร์โบ		
เครื่องชุปเปอร์ชาร์จ	Super Charger	2
เปอร์ชาร์จ		0
ชาร์จเปอร์		0
ชุปเปอร์ชาร์จเจอร์		2
อัดบรรจุอากาศ	Supercharger supercharger	30
อัดบรรจุ		38
บรรจุอากาศ		30
ชุเปอร์ชาร์จ	Super Charge	1

また、ワイルドカードを使用した検索結果を表 2 に示す。ワイルドカードには、記号を挟む文字と文字の間に、0 文字以上の任意の文字を含むことを意味する記号 (%) がある。これを用いることで網羅性を向上させる検索が可能である。

表 2. ワイルドカードの検証

検索語 (supercharger)	検索方法	DIPHTH ヒット件数
อัดบรรจุ	そのまま	38
อัด%บรรจุ	(%)使用	38
อัด%บรรจุ	(%)使用	71

3-3 ベトナム語の検証

ターボチャージャを対象とした検証結果を表 3 に示す。声調記号の有無に関わらず同じ訳があてられた用語を検索すると、それぞれのヒット件数に差がみられた。例えば、TĂNG NAP と TANG NAP の場合、訳はどちらも superchargers だが、それぞれのヒット件数が、前者が 2 件に対して後者が 0 件と相違した。

また、声調記号の有無により異なる訳語を示す場合があり、ヒット件数に差がみられた。例えば、TANG ÁP と TĂNG ÁP の場合、訳は前者が turbochargers、後者が boosters となる。この場合、それぞれのヒット件数が、前者が 0 件に対して後者が 8 件と相違した。

表 3. ベトナム語異表記の検証

ベトナム語	英語 (訳語)	IP LIB ヒット件数
TĂNG NAP	superchargers	2
TẶNG NÁP		0
TANG NAP		0
TANG NÁP		0
TĂNG ÁP	turbochargers	0
TANG ÁP		0
TANG ÁP		0
tăng áp		6
TẶNG ÁP		6
	boosters	6

また、ワイルドカードを使用した検索結果を表 4 に示す。ワイルドカードには、任意の 1 文字を示す記号 (?)、記号を挟む文字と文字の間に 0 文字以上の任意の文字を含むことを意味する記号 (*) がある。しかし、連語検索に用いる記号 (“”) と併用できるのは (*) のみである。従って、(*) 及び (“”) を用いることで網羅性を向上させる検索が可能である。

表 4. ワイルドカードの検証

検索語 (turbochargers)	検索方法	IP LIB ヒット件数
“tăng áp”	そのまま	6
“tăn* áp”	(*) 使用	6
“tăn? áp”	(?) 使用	検索不可
“tă* áp”	(*) 使用	6
“t* áp”	(*) 使用	35

3-4 インドネシア語の検証

ターボチャージャを対象とした検証結果を表 5 に示す。インドネシア語には、声調記号がないため、タイ語、ベトナム語の原語キーワード収集と比較して留意する点が少ない。しかし、スペースの有無によりヒット件数が異なる場合がみられた。また、MESIN TURBO を用いた検索において、IPC が付与されていない公報 (出願番号: P00199700689) をヒットさせることができた。

表 5. インドネシア語異表記の検証

インドネシア語	英語 (訳語)	DGIP ヒット 件数
MESINTURBO	TURBO ENGINE	0
MESIN TURBO		5
TURBOCHARGER	TURBOCHARGER	4
TURBO CHARGER		0
supercharger	supercharger	2
super charger		0

また、ワイルドカードを使用した検索結果を表 6 に示す。ワイルドカードには、任意の 1 文字を示す記号 (_)、記号を挟む文字と文字の間に 0 文字以上の任意の文字を含むことを意味する記号 (%) がある。これらを用いることで網羅性を向上させる検索が可能である。

表 6. ワイルドカードの検証

検索語 (TURBOCHARGER)	検索方法	DGIP ヒット件数
-----------------------	------	---------------

TURBOCHARGER	そのまま	4
T_RBOCHARGER	(_)使用	4
T%RBOCHARGER	(_)使用	4
T%OCHARGER	(%)使用	4
T%CHARGER	(%)使用	13

4. 技術分野ごとの検証

タイ、ベトナム、インドネシアにおけるIPCのみを用いた特許検索の現状を表7に示す。本報告では、日本特許を優先権主張する各国の特許出願を対象とした。そして、日本の特許公報に記載されたIPCを用いてどの程度検索することが可能であるかを検証した。ここでは、上記IPCをそのまま用いて検索できるものを「一致」とした。また、上記IPCのサブクラスまでを用いて検索可能なものを「一部一致」とした。また、上記2つ以外のものを「検索不可」とした。

また、抽出した特許出願に関して、タイについては、1997年から1998年に出願された131件を対象とした。ベトナムについては、2000年から2002年に出願された180件を対象とした。インドネシアについては、2000年から2001年に出願された669件を対象とした(ファミリー抽出にはTotalPatent®を使用)。

表7. IPCのみを用いた特許検索

	TH 件数	VN 件数	ID 件数
一致	56	107	112
一部一致	46	53	77
検索不可	29	20	480
出願件数	131	180	669

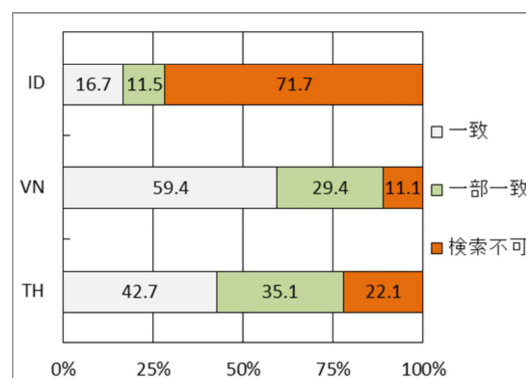


図2. 各分類の比率

表7及び図2より、タイでは29件(22.1%)ベトナムでは20件(11.1%)、インドネシアでは480件(71.7%)の出願が、日本の特許公報に記載されたIPCを用いて検索することが不可能であることがわかった。さらに、検索不可の公報に関して、各国特許庁のデータベースに出願人情報が記載されていないものが、ベトナムで6件、インドネシアで48件みられた。

従って、キーワード検索を併用して網羅性を向上させることは重要である。

5. まとめ

新興国の特許について漏れなく検索するためには、商用データベースのファミリーから対象国の公報情報を得て、原語キーワードを抽出する方法が有効である。そして、網羅性を向上させる検索を行うためには、スペースの有無及びワイルドカードの使用を考慮すべきである。

また、タイ語及びベトナム語の原語キーワード収集に関しては、声調記号の有無を異表記として考慮する必要がある。

さらに、タイ語に関しては、本検証では一方のヒット件数が0件ではあったが、キーワードが連語の場合、語順を変えた異表記にも留意すべきと考えられる。

以上をふまえて、原語が読めない場合であっても、原語キーワードを用いて網羅性を向上させる検索を行う手順案を

以下に示す。

(1)商用データベースから対象国のファミリー情報を得る。

(2)ファミリー情報から、機械翻訳を用いて原語キーワードを抽出する。

(3)原語キーワードについて、さらに機械翻訳を行うことにより、単語の最小構成単位を探索する。加えて、ワイルドカードを用いて異表記を検討する。

(4)原語キーワード及び異表記を用いて検索し、ノイズ判断としてIPC(ただし、付与されている場合)を参照する。

なお、本検証については、発明の名称を対象に行ったが、出願人情報と併用することでIPCを用いた検索を補完し得る、網羅性を向上させる検索が可能と考えられる。

6. おわりに

本報告は、アジア特許情報研究会の2015年度の活動を報告であり、研究会の皆様には情報の提供及び数々の知識や示唆を頂きました。ここに改めてお礼申し上げます。

7. 参考文献

- [1] ジェトロセンサー2012年10月号
http://www.jetro.go.jp/ext_images/jfile/report/07001069/asean_investment.pdf
- [2] 特許行政年次報告書 2015年版
<https://www.jpo.go.jp/shiryoutoushin/nenji/nenpou2015/honpen/1-1.pdf>
- [3] 武藤亜弓 他. “新興国のテクニカルターム検証方法” 第11回情報プロフェッショナルシンポジウム INFOPRO 2014