

## 新興国のテクニカルターム検証方法

### 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

○武藤 亜弓<sup>1)</sup>, 佐野 浩太郎<sup>2)</sup>

株式会社ネットス<sup>1)</sup>, 株式会社IHI<sup>2)</sup>

〒105-0004 東京都港区新橋4丁目5番1号 アーバン新橋ビル3階

Tel: 03-5777-3263 FAX: 03-5777-3264

E-mail: nets-muto@sun-group.co.jp

## Verification Method of Technical Term in Emerging Country Gathering technical term in Thai, Vietnamese and Indonesian

MUTO Ayumi<sup>1)</sup>, SANO Koutaro<sup>2)</sup>

Nets Co., Ltd.<sup>1)</sup>, IHI Corporation<sup>2)</sup>, OLYMPUS Medical Systems Corp.<sup>3)</sup>, Showa Shell  
Sekiyu K.K.<sup>4)</sup>, DENKI KAGAKU KOGYO KABUSHIKI KAISHA<sup>3)</sup>

Urban-Shimbashi Building 3F, 5-1, Shimbashi 4-chome, Minato-ku, Tokyo 105-0004  
JAPAN

Phone: +81-3-5777-3263 Fax: +81-3-5777-3264

E-mail: nets-muto@sun-group.co.jp

### 【発表概要】

新興国の特許は IPC 付与が不十分であり、スクリーニングするための母集団抽出が難しく、IPC による検索を補完するためにはキーワード検索を要する。本研究では英語情報が収録されていない新興国(タイ・ベトナム・インドネシア)で、テクニカルタームがどのように使われているのか調べ、各国の特許調査の精度を高めることを目的とする。また、普段 使い慣れていない言語のテクニカルタームを、効率良く網羅的に収集する方法を、体系化し提案する。

### 【キーワード】

テクニカルターム, 翻訳機能, 表記方法, 特許調査, タイ語, ベトナム語, インドネシア語

## 1. はじめに

近年、日本企業の投資先として、ASEAN 諸国の存在感が高まっているように、市場としても、生産工場としても、ASEAN は日本と密接な関係になっている。特に、タイ、ベトナム、インドネシアの3ヶ国については、2013 年の投資額は、過去最高額を記録する等、日本企業からの注目度は高く、それに伴い、各国における特許調査の必要性も高まっている。

しかしながら、「知財先進国」である日本と比較すると、ASEAN 諸国での特許調査に、日本国内での調査と同じ水準を求めることは不可能に等しい。その理由としては、主に、「特許情報の未整備」と「言語の障壁」という2つの要因が挙げられる。まず、「特許情報の未整備」とは、①知的財産に関連する法制度が整っていない(または、法制度の歴史が浅い)、②特許調査に使用するデータベースが十分に整備されていない(データベースの機能面、収録率等)、③特許公報の情報が不十分(ファミリー付与、IPC付与率が低い等)というように、不十分と一口に言っても、特許情報に関わるあらゆる部分を含んでいる。また、2つ目の問題、「言語の障壁」とは、ASEAN 諸国では、母国語が英語や中国語といった、日本人に比較的馴染みのある言語ではない現地の言語が母国語であることが多く、日本人が、各国の特許についてキーワード検索をしたり、抽出した特許公報を読み込むことが「やりづらい」。このような理由から、ASEAN 諸国での特許調査に、日本の国内調査と同水準を求めることは不可能に等しいが、この特許調査に不利な環境下で、出来る限り網羅的に特許を調査する方法について、検討したいと思う。例えば、ファミリーやIPCを活用する方法も考えられるが、先に述べたように、ASEAN 諸国の特許公報について

は、ファミリー付与やIPC付与率が低いため、ファミリーやIPCのみに頼る方法は危険である。ここで、ファミリーやIPCを使用する検索方法を補完するために、キーワードを使用することが必要となる。しかし、前述の ASEAN での特許調査における2つ目の問題、「言語の障壁」とおり、日本人にとって、キーワード検索に使用するテクニカルタームを収集することが難しい。

このような背景から、本研究では、IPC付与が不十分であり、英語収録がされていない新興国で、テクニカルタームがどのように使われているのかを検証し、テクニカルタームを網羅的・効率的に収集できる方法を体系化し、提案する。

## 2. 3つのアプローチ方法

本研究では、まず、検証するための言葉を、3つのアプローチから収集する。①言語からのアプローチ②分類からのアプローチ③ファミリーからのアプローチ、である。各々の具体的な方法及び、そのメリット・デメリットを以下に述べる。尚、具体例として「半導体」という言葉を挙げて説明する。

### 2-1 言語からのアプローチ

①言語からのアプローチは、以下の手順で言葉を収集する。

- 1.日本語→対象言語に変換(使用ツール:google 翻訳)
- 2.英語→対象言語に変換(使用ツール:google 翻訳)
- 3.1と2で収集した言葉を各国特許庁データベースで、キーワード検索する
- 4.ヒット公報の内容を読み、収集した言葉が実際に特許でも同じ意味で使われているのか確認する(使用ツール:google 翻訳)

この方法は、日本語や英語から、一般的な言葉を検索する方法として優れているが、一方で、使用する翻訳ツールの精

度に頼ってしまい、翻訳ツールで変換される言い方以外の言葉に気付きにくいという欠点がある。また、特許庁のデータベースによっては、キーワードが前方一致でないとヒットしないといった特徴もあるため、①のアプローチ手順1と2で変換した言葉を、特許庁データベースで検索してもヒットしない可能性もあることを留意しなくてはならない。

実際に、このアプローチ方法でタイ語・ベトナム語・インドネシア語の「半導体」について収集すると、表1のようになった。各国特許庁のデータベースにて、抽出した言葉の公報ヒット件数を確認すると、google 翻訳等オンライン辞書で翻訳された言葉が実際に特許でも使われているのか否かがわかる。例えば、「เครื่องสื่อไฟฟ้า」というタイ語の言葉は、公報ヒット件数が0件であった。「特許情報の未整備」という点や表記揺れの問題等を考慮すると、必ずしも「เครื่องสื่อไฟฟ้า」というタイ語が、特許公報で使用されていないとは言い切れないが、オンライン辞書で翻訳されるくらいに比較的一般的に使用されている(と推測される)言葉が、特許でも使用頻度が高いとは限らないことがわかる。一方、ベトナム語やインドネシア語は、日本語からも英語からも、どちらも同じ一通りの言葉しか収集することができなかった。

日本語・ 特許庁	半導體	1	1
英語	semiconductor	1	1
タイ語	เครื่องสื่อไฟฟ้า	0	0
ベトナム語	chips bán dẫn	0	0
インドネシア語	semikonduktor	0	0
英語	semikonduktor	1	1
英語	semikonduktor	1	1
英語	semikonduktor	1	1

日本語・ 特許庁	半導體	1	1
英語	semiconductor	1	1
タイ語	เครื่องสื่อไฟฟ้า	0	0
ベトナム語	chips bán dẫn	0	0
インドネシア語	semikonduktor	0	0
英語	semikonduktor	1	1
英語	semikonduktor	1	1
英語	semikonduktor	1	1

表1 「言語からのアプローチ」により収集した「半導体」に関するタイ語・ベトナム語・インドネシア語リスト

## 2-2 分類からのアプローチ

続いて、②分類からのアプローチは、以下の手順で言葉を収集する。

1. 収集したい言葉の分類を確認
2. 対象国特許庁データベース、または対象国対応の商用データベースで分類を検索
3. ヒット公報の内容を読み、収集した言葉が実際に特許でも同じ意味で使われているのか確認する(使用ツール: google 翻訳)

この方法は、分類からアプローチするという点で、比較的近い技術分野のものを拾うことができる。また、翻訳機能の精度に頼るだけでは見つけられない言葉の発見を期待することができるという点で、①「言語からのアプローチ」を補完できる。一方で、近い技術分野とはいえ、上位の分類で検索するとノイズが増えてしまい、検証しにくい点や、分類が正確に付与されている公報が少ないという点で新興国の特許調査には不向きな点も挙げられる。

表2は、このアプローチ方法で「半導体」について収集したタイ語・ベトナム語・インドネシア語のリストである。今回は、IPC「H01L21」を使用した。網羅性を重視するならば、複数のIPCを使用すべきであるが、検証の時間短縮のため1つのIPCに留める。



ている言葉が、キーワードでヒットしない場合があることが明らかとなった。

また、ファミリーからのアプローチにより、①や②のアプローチ方法で発見できなかった言葉を収集することができた。

### 3. 3つのアプローチ方法による収集結果

前項で述べた3つのアプローチ方法により、以下の表4のように収集することができた。

国名	タイ語	ベトナム語	インドネシア語	英語	日本語
タイ語	เซมิคอนดัคเตอร์			semiconductor	半導体
	ชิป			chip	チップ
	ไอซี			IC	IC
	ไมโครชิป			microchip	マイクロチップ
ベトナム語		chíp		chip	チップ
		chíp		chip	チップ
インドネシア語			chip	chip	チップ
			chip	chip	チップ
英語			chip	chip	チップ
			chip	chip	チップ
日本語			chip	chip	チップ
			chip	chip	チップ

表4 3つのアプローチ方法により収集した「半導体」に関するタイ語・ベトナム語・インドネシア語リスト

表4からわかるとおり、異なる3つのアプローチ方法によって収集できる言葉は同じものもあれば、異なるものもあった。また、特にタイ語の場合、日本語で「半導体」という一つの単語が、複数の表現方法が存在するということが明らかになった。一方ベトナム語については、本検証で抽出できた言葉は1語のみであったが、ベトナム特許庁データベースや商用データベース(PatBase)において、文字化けのため正確な文章で収録されていない公報が複数存在したことから、「bán dẫn」以外にも「半導体」を意味する言葉が存在する可能性も考えられる。

本稿では、タイ語・ベトナム語・インドネ

シア語の「半導体」を例に挙げたが、対象国や技術分野によっても各々のアプローチ方法による収集状況は異なるため、それぞれを補完しあう形で、3つのアプローチ方法を組み合わせる必要があると考えられる。

### 4. まとめ

本研究では、「網羅的に、かつ効率的」にテクニカルタームを収集する方法を見つけることを目的とした。「網羅的」という点では、3つの方法で言語に対してアプローチすることにより、出来るだけ漏れの少ないキーワード検索が実現できると考えられる。一方、「効率的」という点では、3つの方法を組み合わせるという点で、既に、3つのアプローチのうち一つだけを選択する方法に比べると、時間がかかる作業ではあるが、ただやみくもにキーワードを収集する方法に比べると、3つの方法を組み合わせることによって網羅性を実現できるという点で「効率的」と言えよう。

ただし、②分類からのアプローチについては、「どのIPC分類を使用するか」といった点について、調査をする側の裁量に大きく委ねられているため、議論の余地がある。また、③ファミリーからのアプローチについても、本研究では、PatBaseを使用したか、どの商用DBを使うかによって、ファミリーの収録状況や機能が異なるため、網羅性や作業効率に変化があることも留意したい。

### 5. おわりに

本報告はアジア特許情報研究会の2014年度の活動を報告するものである。

### 6. 参考文献

[1]「ASEAN をめぐる動向」ジェトロ世界貿易投資報告 2014 年度版 日本貿易振興機構(2014年9月26日)

<http://www.jetro.go.jp/world/gtir/2014/pdf/2014-asean.pdf>