

2021年3月25日

データベース収録確認ワード(その2)

アジア特許情報研究会:伊藤徹男

1. はじめに

ワールドワイドに各国をカバーしているデータベース、特に商用データベースでは収録国数の多さと古い年代から収録していることに加え、データベース全体の収録数を誇らしげに示されることが多いようですが、調査担当者の立場からは調査対象国のデータが100%収録され(対象国のデータベースと同じ収録)、書誌情報はもちろん、クレームや全文も検索対象になっているかどうかデータベース選択の重要なポイントだと思っています。

東アジア諸国のデータは英語で提供されるものが多くなりましたが、原語も収録されるようになり、英語+原語のハイブリッド検索が可能となっています。ASEAN 各国のデータはまだ原語のままのものもありますが、5年ほど前の PATENTSCOPE には原語も収録されていませんでしたので、それに比べれば収録しているだけでもありがたいです。

要約情報の収録が一部だけであるのに気がつかず、調査した結果、目的の情報が得られずに「調査しましたがありませんでした」と報告するのは悲しいことです。報告を受ける側はデータベースの収録状態など「そんな細かなことまで把握できない」のです。

収録が不完全な状態を認識して調査するのであれば(そのような前提も報告して)、まだ幸いです。収録の把握は調査担当者の責任です。

前回は、2009年に検証した「データベース収録確認ワード」として東アジア3国データベースの収録を英語と原語から確認していた方法を紹介すると共に、その手法が最近のワールドワイドなデータベースについて適用可能かどうかを英語ワードで再検証しました。

本稿では、さらに東アジア3国原語の収録確認ワードが各種データベースに適用可能かどうかを確認したいと思います。

商用英語データベースでさえ、原語から英語への翻訳の過程で誤訳したり(外国からの出願では英語情報から原語に翻訳する際にも誤訳が生じます)、英語のスペルミスのために「英語情報だけを検索しては検索洩れを生じる」ことについては、これまでもあちこちで指摘してきました。

英語+原語情報からハイブリッド検索で洩れを如何に防ぐことができるか、を担保するためにも原語情報の収録がどの程度かも把握することは重要です。

2. データベース収録確認ワードの再検証

ここでは前回の英語収録確認ワードの検証に引き続き、原語ワード集合について議論します。

2009年時点では、原語データベースの収録確認ワードとして以下のものを使っていました。

中国大陸特許収録確認ワード

2007年CN公開特許:208,345件

AB (種 or 一 or 发 or 的 or 用 or 在) 208,332件(≒100%)

CL (種 or 一 or 1 or 的 or 中 or 于 or 在) 208,325件(≒100%)

FULL (種 or 一 or 1 or 的 or 用 or 和 or 中 or 上 or 于 or 在) 208,331件(≒100%)

台湾特許収録確認ワード

2007年公開特許 46986件／公告特許 22218件

AB 一 or 種 or 發 or 的 or 用 or 以 or 含 or 具 or 其 or 之 or 本 or 。

⇒ 公開 46981件(≒100%)／公告 22218件(=100%)

CL 一 or 1 or 種 or 的 or 用 or 以 or 下 or 含 or 具 or 其 or 之 or 。

⇒ 公開 46759件(≒100%)／公告 22218件(=100%)

FULL 一 or 1 or 種 or 發 or 的 or 用 or 以 or 下 or 具 or 其 or 之 or 本 or 。

⇒ 公開 46766件(≒100%)／公告 22214件(≒100%)

句読点も収録確認ワードとして使えることもわかりました。

韓国特許特許収録確認ワード

1+2+3+1+.....+.....+.....+.....+

2007年公開特許 122581件

AB 122428件／CL 116759件／FULL 122552件

1) 原語収録確認ワードの再検証

上記原語収録確認ワードは参考情報として、2020年発行公開特許数を基に中国、台湾、韓国データベースで出現頻度の高い用語を試行錯誤的に抽出し、発行日書誌を100とした出現頻度を求めました。Espacenet のように検索タームの入力制限がなければ確認用語は多いほど収録率は高くなりますが、ワード集合に追加しても収録率に変化ない場合にはできるだけ少ない用語数としました。

中国、台湾、韓国ともいずれも1語で検索できますし、2007年の台湾検証では句読点からも収録を確認できましたので3国とも句読点や出現頻度の高そうな記号なども加えて検証しました。

その結果、以下のワード集合を収録確認ワードとすることとしました。(算出根拠:表1～表3)

① 中国 2020年公開特許(1517113件): 表1

AB:的* or 一* or 本* or 种* or 发* or 明* or 用* or 和* or 于* or 及* or 有* or 在* or 中* or 以* or 所* or 内*(1517108件(≒100%))

CL:的* or 其* or 所* or 一* or 述* or 在* or 于* or 种* or 中* or 和* or 有* or 用* or 上* or 以* or 内*(1517013 件(≒100%))

FULL:的* or 本* or 一* or 在* or 所* or 种* or 中* or 有* or 和* or 于* or 述* or 上* or 其* or 用* or 内*(1517017 件(≒100%))

② 台湾 2020年公開特許(46829 件) 表2

AB:。 or 一 or 的 or 種 or 以 or 之 or 本(46824 件(≒100%))

CL:l or 。 or 一 or 種 or 其 or 以 or : or 、 or 之(46827 件(≒100%))

FULL:。 or 一 or 本 or 以 or 用(46828 件(≒100%))

③ 韓国 2020年公開特許(146030 件) 表3

AB:** or * or ** or * or * or ** or * or ** or * or **
(145972 件(≒100%))

CL:** or * or ** or * or * or ** or l* or * or * or **
(145726 件(≒100%))

FULL:** or * or ** or * or l* or * or * or * or * or **
(146027 件(≒100%))

表1. 中国公開特許収録確認ワード出現率(CNIPR 要約出現率でソート)

	TI		AB		CL		FULL	
的	705,454	46%	1,508,207	99%	1,511,258	100%	1,512,499	100%
一	948,093	62%	1,450,614	96%	1,502,015	99%	1,510,583	100%
本	8,601	1%	1,426,223	94%	218,603	14%	1,512,061	100%
种	942,706	62%	1,381,009	91%	1,445,099	95%	1,503,024	99%
发	45,055	3%	1,331,717	88%	460,151	30%	146,142	10%
明	6,185	0%	1,277,055	84%	93,904	6%	49,503	3%
用	324,613	21%	1,089,991	72%	1,183,697	78%	1,392,714	92%
和	173,653	11%	1,065,699	70%	1,273,570	84%	1,494,701	99%
于	241,395	16%	1,058,710	70%	1,494,517	99%	1,453,268	96%
及	560,222	37%	1,025,619	68%	766,064	50%	1,239,014	82%
有	61,028	4%	1,016,874	67%	1,201,473	79%	1,494,846	99%
在	23,969	2%	970,816	64%	1,496,820	99%	1,510,396	100%
中	56,881	4%	887,508	58%	1,296,692	85%	1,499,614	99%
以	34,915	2%	848,623	56%	1,032,947	68%	1,325,667	87%
所	2,086	0%	797,251	53%	1,502,755	99%	1,505,846	99%
法	764,904	50%	784,683	52%	866,722	57%	301,944	20%
上	16,909	1%	744,143	49%	1,136,335	75%	1,438,321	95%

:	0	0	0	0
。	0	0	0	0
、	0	0	0	0

中国特許データベース中には句読点として句点「。」、読点「、」を含みます。CNIPR では検索できませんが、PATENTSCOPE や TWPAT、Global Patent Search System(GPSS)では収録率100%として検索できますが、収録確認ワード集合には採用していない。

図1. CNIPR 要約中の句点、読点(検索不可)

摘要：本发明公开了一种苹果种子盆栽种植方法，所述苹果种子盆栽种植方法，步骤一：选用十枚苹果种子和催芽盘，步骤二：首先是在吃苹果的时候，小心的把种子完好的拿出来，轻轻把苹果种子上的果肉刮干净，然后把种子泡在水中8-12小时，步骤三：然后把种子放在湿的餐巾纸上，可以盖上塑料膜在25度的环境下2-3天对苹果种子进行催芽，步骤四：也可以不进行催芽操作，直接拿出一个小花盆，用工具挖出几个小坑，每个坑中放一粒苹果种子。每个坑注意只放一颗就好，不要放太多。与现有技术相比，本发明采用盆景方式进行培育，不仅可以观赏，而且所挂的果实也可以食用，大大的提高了苹果培育的经济效益，拓宽了苹果树的种植方式和销售。机器翻译

図2. PATENTSCOPE 要約中の句点、読点(検索可)

[ZH]
一种作为放射线检查装置的X射线检查装置(60)，具备：X射线源(71)，其是对被检查物照射作为放射线的X射线的放射线源；X射线传感部(72)，其是接受来自X射线源(71)的放射线的放射线传感部；以及放射线遮蔽部(82)，其在X射线传感部(72)侧延伸于放射线直接入射的照射区域之外，遮蔽在X射线传感部(72)散射的放射线。由此，能够降低由放射线在放射线传感部的散射引起的放射线的泄漏。

PATENTSCOPE で原語要約が欠落している場合は要約部分が空白か以下のように収録待ちのアナウンスがあります。

図3. PATENTSCOPE で原語要約が欠落している例

Abstract

[EN]

The invention discloses a food with the effects of clearing and nourishing the throat. The raw materials include pear and fritillaria and the adjuvant material is sugar. In order to solve the problems of bitter tastes and colors of the Chinese medicinal materials, in the invention, the residues after extracting the pear are mixed with the traditional Chinese medicine and then the mixture is soaked, thus better solving the problems of bitter tastes and colors, and then the residues are extracted and various active ingredients are preserved, thus reducing the usage of sugar, reducing the irritation of sugar to the throat and enhancing the pharmacological effects.

[ZH]

局待查。

表2. 台湾公開特許収録確認ワード出現率(TWPAT 要約出現率でソート)

	TI		AB		CL		FULL	
。	11	0%	46797	100%	46820	100%	46826	100%
一	1247	3%	42144	90%	46810	100%	46824	100%
的	12369	26%	38163	81%	41905	89%	46632	100%
種	916	2%	36617	78%	46690	100%	46672	100%
以	3198	7%	33787	72%	43826	94%	46819	100%
之	10605	23%	31505	67%	42761	91%	46686	100%
本	135	0%	29506	63%	5762	12%	46821	100%
發	1692	4%	27375	58%	12537	27%	46784	100%
其	8271	18%	25839	55%	46413	99%	46761	100%
用	9216	20%	25815	55%	32853	70%	46804	100%
、	6882	15%	25517	54%	34446	74%	46722	100%
具	3561	8%	22333	48%	32875	70%	46433	99%
舍	1327	3%	20760	44%	30472	65%	41124	88%
:	2	0%	17099	37%	43487	93%	43008	92%
下	232	0%	11745	25%	27587	59%	46466	99%
1	277	1%	8389	18%	46827	100%	46335	99%
(半)	0	0%	0	0%	0	0%	0	0%

いずれも読点(.)だけではほぼ100%を示しますが。

表3-1. 韓国公開特許収録確認ワード出現率 (WIPS GLOBAL 要約出現率でソート)

	TI		AB		CL		FULL	
하*	26586	18%	140243	96%	142403	98%	146007	100%
이*	35338	24%	116070	79%	90586	62%	145990	100%
및*	69161	47%	113495	78%	112391	77%	143404	98%
의*	12542	9%	107389	74%	95074	65%	145681	100%
제*	35963	25%	105449	72%	86013	59%	145272	99%
본*	272	0%	104634	72%	8308	6%	145553	100%
발*	6235	4%	99856	68%	20787	14%	145019	99%
한*	17905	12%	88565	61%	40315	28%	145828	100%
기*	25703	18%	82667	57%	77147	53%	145479	100%
있*	1052	1%	82649	57%	52904	36%	145948	100%
1*	1121	1%	62293	43%	77081	53%	145726	100%
전*	18340	13%	58037	40%	56037	38%	142647	98%
배*	4087	3%	35379	24%	37224	25%	127431	87%
에*	3889	3%	28780	20%	28788	20%	138725	95%
설*	1904	1%	27553	19%	26880	18%	144507	99%
(半)	572	0%	145943	100%	144946	99%	146029	100%
(半)	17698	12%	141872	97%	137209	94%	145948	100%
。	0	0%	0	0%	0	0%	23	0%

半角ドット、半角カンマのみで高い収録率を示したが、収録確認ワードとして採用していません。

表3-2. 韓国公開特許収録確認ワード出現率 (PATENTSCOPE 要約出現率でソート)

	TI		AB		CL		FULL	
하	34239	18%	182928	98%	182946	98%	184544	99%
에	17933	10%	182405	97%	182161	97%	184252	98%
의	54678	29%	178018	95%	183320	98%	184898	99%
한	47848	26%	174966	93%	158995	85%	185283	99%
을	32181	17%	174358	93%	181785	97%	184784	99%
를	40936	22%	172780	92%	179210	96%	184497	99%
이*	50835	27%	166202	89%	179923	96%	187267	100%
및*	83684	45%	145075	77%	165792	89%	184545	99%
분	313	0%	141710	76%	22932	12%	186740	100%
제*	46739	25%	139954	75%	176423	94%	186500	100%
발	9290	5%	137796	74%	57084	30%	184217	98%
로*	7150	4%	112336	60%	166538	89%	186753	100%
있*	1327	1%	110515	59%	177997	95%	187207	100%
기*	28215	15%	109185	58%	143026	76%	187281	100%
1*	1272	1%	78239	42%	187282	100%	187282	100%
전*	20713	11%	73777	39%	115500	62%	182938	98%
태	6846	4%	49179	26%	82880	44%	182750	98%
설	3993	2%	43366	23%	72801	39%	182801	98%
정	1604	1%	8676	5%	182787	98%	137969	74%

(半) 187287 100% 187287 100% 187287 100% 187287 100%

。 187287 100% 187287 100% 187287 100% 187287 100%

データベースが異なると若干、収録確認ワード集合にも差があるようです。

2) データベースの原語収録状況

上記検証で得られた東アジアの原語ワード集合を用いて PATENTSCOPE および Global Patent Search System(GPSS)などワールドワイドな国を収録しているデータベースの要約、クレーム等の収録状況を確認しました。原語収録が始まった Espacenet についてもどの程度原語が収録されているのか興味あるところですが、現段階では出願人以外に原語検索できないことは前回紹介しました。用語について原語検索可能な状況になったら確認してみたいと思います。

PATENTSCOPE には台湾特許の収録がありませんので中国と韓国について、台湾特許庁の GPSS については自国台湾の発明の名称～詳細な説明までは TWPAT 同様、繁体字は100%収録されていると思いますが、収録確認ワード集合の妥当性を確認するためにも中国と韓国と共に調べました。いずれも公開特許の収録状況を確認することで各データベースの収録状況を推察することにしました。

英語商用データベースに東アジアの登録特許収録が始まった2010年頃(この時点では原語収録している商用データベースはなし)にはその状況が気になって調べたことがありますが、現段階では東アジア3国についてはほぼ英語登録特許は収録されているものと思います。権利侵害調査では登録特許の収録状況が重要ですが、次回以降、登録特許の原語収録も確認します。

① PATENTSCOPE の中国、韓国公開特許収録状況

中国特許

英語収録状況では2012～2018年の収録が不完全でしたが、書誌情報そのものが2013年以降不十分な状態です。中国語の書誌に対する要約、クレーム、詳細な説明の収録はほぼ問題ありません。データ更新により書誌情報と共に要約等の収録も充実していくものと思います。

表4. PATENTSCOPE 中国特許原語収録

PD	CNIPR		PATENTSCOPE					
	CN公開	CN公開	AB		CL		DE	
2000	38,296	38,296	38295	100%	38293	100%	38295	100%
2001	50,364	50,365	50353	100%	50358	100%	50363	100%
2002	58,984	58,984	58977	100%	58971	100%	58981	100%
2003	77,472	77,472	77471	100%	77458	100%	77469	100%
2004	93,944	93,943	93943	100%	93933	100%	93943	100%
2005	155,446	155,445	155444	100%	155425	100%	155440	100%
2006	172,424	172,428	172425	100%	172399	100%	172407	100%
2007	208,345	208,348	208341	100%	208311	100%	208327	100%
2008	241,182	241,183	241178	100%	241159	100%	241181	100%
2009	281,006	281,005	281002	100%	280982	100%	280991	100%
2010	315,836	315,844	315839	100%	315807	100%	315830	100%
2011	368,434	368,452	368448	100%	356519	97%	356535	97%
2012	543,296	542,959	539112	99%	537973	99%	538009	99%
2013	632,585	594,187	591351	100%	590142	99%	590162	99%
2014	777,335	751,525	751140	100%	747557	99%	747590	99%
2015	955,341	845,387	845255	100%	833072	99%	833088	99%
2016	1,045,796	937,445	902667	96%	920726	98%	920742	98%
2017	1,270,655	1,094,814	1036481	95%	1058349	97%	1058393	97%
2018	1,575,621	1,595,924	1595919	100%	1593035	100%	1593127	100%
2019	1,532,682	1,543,440	1543433	100%	1543186	100%	1543191	100%
2020	1,517,110	1,428,144	1428135	100%	1428123	100%	1428127	100%

参考までに1985～1999年の公開特許収録をざっくり比較すると以下ようになります。

書誌:(CNIPR 233348/PATENTSCOPE 163406(70%))

PATENTSCOPE 書誌に対し、AB:125963(77%)、CL:125952(77%)、DE:125970(77%)です。

どの年代が欠落しているかなども次回報告します。

韓国特許

KIPRISとPATENTSCOPEの書誌情報乖離については未検証です。英語収録状況は2011年以降の要約収録が不完全でしたが、2012～2017年のクレーム、詳細な説明の収録が若干不完全な点を除けば問題ないと言えます。

表5. PATENTSCOPE 韓国特許原語収録

PD	KIPRIS		PATENTSCOPE					
	KR公開	KR公開	AB		CL		DE	
2000	77,495	97,283	96965	100%	93811	96%	93811	96%
2001	114,264	122,774	122470	100%	121605	99%	121605	99%
2002	97,487	103,539	103058	100%	102749	99%	102749	99%
2003	97,909	101,973	101084	99%	101946	100%	101946	100%
2004	110,582	115,168	114958	100%	115121	100%	115121	100%
2005	123,492	130,240	129971	100%	130230	100%	130230	100%
2006	135,946	157,370	156908	100%	157358	100%	157358	100%
2007	122,581	162,915	162087	99%	162910	100%	162911	100%
2008	122,581	146,537	120720	82%	146528	100%	146534	100%
2009	133,121	147,622	137899	93%	147621	100%	147621	100%
2010	139,151	152,886	129393	85%	152885	100%	152885	100%
2011	140,131	157,823	157621	100%	157820	100%	157820	100%
2012	139,463	159,842	159630	100%	91032	57%	91032	57%
2013	141,194	169,094	168922	100%	130015	77%	130015	77%
2014	148,110	185,559	185253	100%	138651	75%	138651	75%
2015	146,108	178,771	176895	99%	166019	93%	166019	93%
2016	150,599	177,762	169521	95%	156250	88%	156250	88%
2017	143,454	180,263	179332	99%	162927	90%	162927	90%
2018	138,541	175,567	175545	100%	173787	99%	173787	99%
2019	143,824	182,115	182084	100%	182114	100%	182114	100%
2020	146,008	187,287	187254	100%	187287	100%	187287	100%

中国同様、1980～1999年の韓国公開特許収録をざっくり比較すると以下のようになります。

書誌: (PATENTSCOPE 530838)

PATENTSCOPE 書誌に対し、AB:338764(64%)、CL:234784(44%)、DE:234784(44%)です。

やはり古い年代は収録が不十分なようです。

② GPSSの中国、台湾、韓国公開特許収録状況

中国特許

GPSSの中国特許では、2000年以前の「詳細な説明」の収録が(現時点では)ありませんが、全体に素晴らしい収録状況です。前にも紹介したようにGPSS(TWPAT共に)では中国と台湾特許

は簡体字でも繁体字でも検索・表示できるようになっていますのでその点でも使いやすいです。

表6. GPSS 中国特許原語収録

PD	CNIPR		GPSS					
	CN公開	CN公開	AB		CL		DE	
2000	38,296	38,296	38295	100%	38209	100%	0	0%
2001	50,364	50,365	50354	100%	50357	100%	50360	100%
2002	58,984	58,984	58979	100%	58970	100%	58980	100%
2003	77,472	77,472	77471	100%	77462	100%	77470	100%
2004	93,944	93,944	93943	100%	93933	100%	93942	100%
2005	155,446	155,447	155447	100%	155426	100%	155440	100%
2006	172,424	172,428	172425	100%	172398	100%	172405	100%
2007	208,345	208,348	208341	100%	208288	100%	208298	100%
2008	241,182	241,182	241178	100%	241163	100%	241182	100%
2009	281,006	281,007	281005	100%	280998	100%	281006	100%
2010	315,836	315,840	315833	100%	315814	100%	315836	100%
2011	368,434	368,434	368433	100%	368420	100%	368430	100%
2012	543,296	543,297	543296	100%	543264	100%	543296	100%
2013	632,585	632,585	632560	100%	632565	100%	632585	100%
2014	777,335	777,336	777331	100%	777304	100%	777336	100%
2015	955,341	955,341	954993	100%	955324	100%	955341	100%
2016	1,045,796	1,045,741	1045734	100%	1045425	100%	1045451	100%
2017	1,270,655	1,270,363	1270343	100%	1270303	100%	1270363	100%
2018	1,575,621	1,575,438	1575430	100%	1575342	100%	1575436	100%
2019	1,532,682	1,532,539	1532538	100%	1532525	100%	1532530	100%
2020	1,517,110	1,517,011	1517008	100%	1517004	100%	1517008	100%

台湾特許

GPSS は台湾特許庁データベースであり、原語(繁体字)収録は台湾データのみを収録している従前から存在している TWPAT と同一であるので繁体字での収録確認も無意味なようであるが、新たに検証した原語収録確認ワードの妥当性を確認する点で敢えて掲載しておくことにしました。

一部 100 件以上食い違いのある年代もあるが、ほぼ問題ないと判断した。さらなる検討は後日実施したいと思っています。

表7. GPSS 中国特許原語収録

PD	TW公開	GPSS					
		AB		CL		DE	
2003	8,194	8191	100%	8185	100%	8190	100%
2004	28,927	28925	100%	28916	100%	28919	100%
2005	41,439	41438	100%	41422	100%	41425	100%
2006	44,783	44778	100%	44623	100%	44628	100%
2007	46,986	46981	100%	46759	100%	46765	100%
2008	50,141	50139	100%	49939	100%	49940	100%
2009	52,618	52618	100%	52607	100%	52607	100%
2010	44,962	44962	100%	44960	100%	44951	100%
2011	46,157	46157	100%	46156	100%	46146	100%
2012	51,592	51591	100%	51592	100%	51592	100%
2013	52,126	52124	100%	52126	100%	52125	100%
2014	48,719	48719	100%	48717	100%	48719	100%
2015	47,367	47366	100%	47367	100%	47367	100%
2016	44,356	44355	100%	44355	100%	44356	100%
2017	43,677	43673	100%	43671	100%	43677	100%
2018	44,071	44070	100%	44060	100%	44070	100%
2019	47,987	47986	100%	47985	100%	47986	100%
2020	46,829	46828	100%	46827	100%	46828	100%

韓国特許

GPSS の韓国原語収録は表8を見てわかるように、まったく使い物になりません。クレームや詳細な説明は英語情報も未収録です。

表8. GPSS 韓国特許原語収録

PD	GPSS		
	KR公開	AB	
2000	98966	73538	74%
2001	116634	80021	69%
2002	97510	97236	100%
2003	97917	76358	78%
2004	110593	41574	38%
2005	123502	258	0%
2006	136148	434	0%
2007	123910	1477	1%
2008	121568	6825	6%
2009	146383	13555	9%
2010	152270	14141	9%
2011	152189	12407	8%
2012	148473	9451	6%
2013	148666	8488	6%
2014	154602	7998	5%
2015	150775	100795	67%
2016	155142	155059	100%
2017	147879	147835	100%
2018	142232	142181	100%
2019	147105	147046	100%
2020	148875	148832	100%

③ WIPS GLOBAL の中国、韓国公開特許収録状況

WIPS GLOBAL は韓国の WIPS 社が提供するワールドワイドな商用特許データベースですが、韓国特許庁(韓国特許情報院)の KIPRIS では複合検索などで若干使いにくいなどもあり、韓国特許調査だけでなく、最近では中国、台湾特許調査にも利用しています。

韓国のベンダーさんなので韓国特許の原語収録は問題ないとは思いますが、英語情報の収録と併せて確認しました。ここでも確認は公開特許のみです。

中国特許

英語も原語(簡体字中国語)も問題なく収録されているようです。表中の「DSC」は詳細な説明部分からの抽出用検索コマンドです。

表9. WIPS GLOBAL 中国特許英語および原語(簡体字)収録

PD	英語収録						簡体字収録						
	CN書誌	AB		CL		DSC		AB		CL		DSC	
2000	38296	38290	100%	38244	100%	38258	100%	38295	100%	38294	100%	38296	100%
2001	50364	50351	100%	50317	100%	50330	100%	50353	100%	50361	100%	50364	100%
2002	58984	58975	100%	58921	100%	58936	100%	58979	100%	58974	100%	58984	100%
2003	77471	77449	100%	77361	100%	77382	100%	77470	100%	77462	100%	77470	100%
2004	93944	93926	100%	93813	100%	93835	100%	93943	100%	93935	100%	93944	100%
2005	155446	155425	100%	155289	100%	155318	100%	155445	100%	155432	100%	155446	100%
2006	172424	172406	100%	172208	100%	172253	100%	172423	100%	172417	100%	172424	100%
2007	208345	208329	100%	208061	100%	208121	100%	208343	100%	208335	100%	208345	100%
2008	241182	241170	100%	240346	100%	240400	100%	241179	100%	241163	100%	241182	100%
2009	281006	280998	100%	280586	100%	280640	100%	281005	100%	280998	100%	281006	100%
2010	315836	315825	100%	315380	100%	315455	100%	315834	100%	315814	100%	315836	100%
2011	368434	368427	100%	367674	100%	367985	100%	368433	100%	368420	100%	368434	100%
2012	543296	543285	100%	542102	100%	542230	100%	543295	100%	543264	100%	543296	100%
2013	632585	632574	100%	631392	100%	631545	100%	632560	100%	632565	100%	632585	100%
2014	777335	777320	100%	775559	100%	775744	100%	777330	100%	777303	100%	777335	100%
2015	955341	955334	100%	953021	100%	953240	100%	955324	100%	955324	100%	955341	100%
2016	1045718	1045704	100%	1044067	100%	1044260	100%	1045714	100%	1045691	100%	1045718	100%
2017	1270363	1270358	100%	1270155	100%	1270363	100%	1270355	100%	1270303	100%	1270363	100%
2018	1575438	1575426	100%	1575245	100%	1575438	100%	1575433	100%	1575343	100%	1575438	100%
2019	1532539	1532528	100%	1532072	100%	1532539	100%	1532538	100%	1532534	100%	1532539	100%
2020	1517011	1517005	100%	1516932	100%	1517011	100%	1517008	100%	1517007	100%	1517011	100%

台湾特許

一時、台湾特許の「詳細な説明」部分が未収録でしたが、表 10 を見る限りは問題なさそうです。

表10. WIPS GLOBAL 台湾特許英語および原語(繁体字)収録

PD	英語収録						繁体字収録						
	TW書誌	AB		CL		DSC		AB		CL		DSC	
2003	8194	8192	100%	8174	100%	8179	100%	8190	100%	8182	100%	8189	100%
2004	28927	28900	100%	28887	100%	28894	100%	28922	100%	28890	100%	28897	100%
2005	41438	41093	99%	41390	100%	41393	100%	41433	100%	41400	100%	41396	100%
2006	44783	44619	100%	44611	100%	44660	100%	44778	100%	44609	100%	44661	100%
2007	46986	46825	100%	46606	99%	46624	99%	46979	100%	46613	99%	46624	99%
2008	50141	50077	100%	49894	100%	49902	100%	50133	100%	49906	100%	49902	100%
2009	52618	52519	100%	52424	100%	52421	100%	52617	100%	52433	100%	52423	100%
2010	44962	44916	100%	44918	100%	44910	100%	44959	100%	44921	100%	44913	100%
2011	46157	46105	100%	46120	100%	46097	100%	46157	100%	46129	100%	46100	100%
2012	51592	51543	100%	51558	100%	51554	100%	51588	100%	51567	100%	51555	100%
2013	52126	52085	100%	49802	96%	49797	96%	52123	100%	49813	96%	49799	96%
2014	48720	48702	100%	45803	94%	45787	94%	48720	100%	45813	94%	45791	94%
2015	47367	47365	100%	47248	100%	47280	100%	47267	100%	47271	100%	47263	100%
2016	44356	44352	100%	44163	100%	44262	100%	44345	100%	44268	100%	44267	100%
2017	43677	44069	101%	43561	100%	43622	100%	43672	100%	43656	100%	43636	100%
2018	44071	47983	109%	43885	100%	44009	100%	44064	100%	44062	100%	44034	100%
2019	47987	47983	100%	47961	100%	47908	100%	47977	100%	47969	100%	47942	100%
2020	46828	46824	100%	46647	100%	46756	100%	46826	100%	46803	100%	46790	100%

韓国特許

原語(ハングル)収録は当然のこととして、英語情報も問題なく収録していることを確認できました。

表11. WIPS GLOBAL 韓国特許英語および原語(ハングル)収録

PD	英語収録						ハンダール収録						
	KR書誌	AB		CL		DSC		AB		CL		DSC	
2000	77504	77504	100%	77439	100%	77504	100%	77486	100%	77431	100%	77504	100%
2001	114269	114264	100%	114059	100%	114267	100%	114237	100%	114146	100%	114269	100%
2002	97487	97482	100%	97421	100%	97486	100%	97459	100%	97383	100%	97487	100%
2003	97909	97907	100%	97852	100%	97909	100%	97873	100%	97806	100%	97909	100%
2004	110584	110584	100%	110530	100%	110584	100%	110543	100%	110455	100%	110584	100%
2005	123498	123495	100%	123423	100%	123497	100%	123482	100%	123391	100%	123497	100%
2006	135952	135949	100%	135879	100%	135948	100%	135930	100%	135827	100%	135950	100%
2007	122581	122578	100%	122500	100%	122580	100%	122537	100%	122456	100%	122581	100%
2008	114828	114823	100%	114711	100%	114827	100%	114774	100%	114700	100%	114828	100%
2009	133121	133117	100%	132987	100%	133117	100%	133016	100%	132970	100%	133117	100%
2010	139151	139146	100%	138951	100%	139141	100%	138909	100%	138891	100%	139139	100%
2011	140131	140124	100%	139997	100%	140127	100%	140027	100%	139937	100%	140128	100%
2012	139463	139457	100%	139321	100%	139463	100%	139266	100%	139298	100%	139463	100%
2013	141194	141186	100%	141109	100%	141189	100%	141066	100%	141027	100%	141190	100%
2014	148110	148097	100%	148025	100%	148106	100%	148003	100%	147964	100%	148107	100%
2015	146108	146079	100%	145911	100%	146086	100%	146003	100%	145901	100%	146088	100%
2016	150599	150581	100%	150426	100%	150593	100%	150532	100%	150428	100%	150591	100%
2017	143454	143447	100%	143310	100%	143453	100%	143422	100%	143249	100%	143452	100%
2018	138542	281977	204%	138386	100%	138533	100%	138519	100%	138371	100%	138538	100%
2019	143825	143814	100%	143655	100%	143823	100%	143799	100%	143677	100%	143822	100%
2020	146030	145992	100%	145872	100%	146028	100%	145999	100%	145825	100%	146029	100%

④ Espacenet の東アジア原語情報

Espacenet にも 2019 年から中国特許情報が、2020 年から韓国特許情報が原語でも収録され始めました。現時点では中国や韓国の原語が要約やクレームにも部分的に見られるようになりましたが原語から検索できるのは出願人のみです。

台湾特許情報は PATENTSCOPE には収録されていませんが、Espacenet には収録されています。中国、台湾、韓国の東アジア特許情報が原語からも検索できる日も近いでしょう。

図4のようなケースでは、要約中から英語でも原語(簡体字)でも抽出できません。

図4. 要約は原語のみで英訳要約がない。

CNB(CN110765673B)

Abstract

本发明基于等效源法的噪声声场重建方法，包括：对全息面声压信号进行傅里叶变换，求出全息面点阵声压列向量P中每一向量的幅值与相位；求出全息面与等效源间的传递矩阵G1；利用正则化方法求解G1的正则化参数；比较等效源的个数与全息面的点阵个数的大小，若等效源的个数大于全息面的点阵个数则进入S5，否则进入S6；S5、欠定，采用循环迭代求逆法求解等效源权重系数矩阵并进入S7；S6、超定，利用正则化参数求解G1逆矩阵，利用Q=G1-1P公式求解等效源权重系数矩阵Q，并进入S7；求出声源面与等效源之间的传递矩阵G2；S8、利用P0=QG2求得声源面的声压分布实现噪声声场重建。

TWB(TWI690923B)

Abstract

一種記憶體，包含：記憶胞陣列、複數條匹配線和複數對搜尋線。記憶胞陣列包含複數個記憶胞。記憶胞之任一者包含輸出端、二個整流元件和二個電阻元件。二個電阻元件經配置以儲存代表資料狀態之二個位元值。匹配線分別耦接記憶胞的輸出端。搜尋線分別耦接記憶胞。複數對搜尋線之任一者包含第一搜尋線和第二搜尋線。同一個記憶胞的第一電阻元件和第一整流元件串聯於同一對搜尋線的第一搜尋線和此記憶胞的輸出端之間。同一個記憶胞的第二電阻元件和第二整流元件串聯於同一對搜尋線的第二搜尋線和此記憶胞的該輸出端之間。

図5. 英語要約も未収録である、との表示

No abstract found. Please consult other publications of this patent family in "Available in", if displayed above.

3. 今後の検証予定

原語収録状況をそれぞれのデータベースで具体的な検索ターム(用語)を基に検索比較してみたいと思っています。

CN:CNIPR/GPSS/PATENTSCOPE/WIPS GLOBAL

TW:GPSS/WIPS GLOBAL

KR:KIPRIS/GPSS/PATENTSCOPE/ WIPS GLOBAL

4. おわりに

「何故、無料データベースにこだわるのか？」

「実務では無料データベースなんか使いません」というのは調査環境に恵まれた大手企業の調査担当者のみです。知財に関心ある企業、研究機関、大学、個人などで有料の商用データベースを利用できる環境にあるのは極めてわずかな企業、団体ではないかと思います。知財情報を必要としながら利用頻度も少ないため、わずか年間10万円程度のデータベース利用料金さえ払えない知財関係者も存在するのです。

ある特許事務所の出願も調査も一人で担当しているという方から「最近はお願依頼が激減し、調査まで自分でやらなければいけなくなった。外国特許調査ではどんなデータベースを使えばよいか」と質問され、しかも今まで商用データベースなどは使ったことがない、というのは驚きでした。特許事務所と銘打っているところであれば外国特許調査ができるデータベースの1つぐらい・・・、とっていました。自分がいかに恵まれた環境にあったかを思い知らされました。業務で無料データベースを駆使しているところもあるんだ、と。

「特許調査の基本」

特許調査ではデータベースを利用して調査しますが、そのデータベースで最も重要なのは対象

国の情報が 100%収録されているかどうかです。しかも書誌情報だけでなく、少なくとも要約(抄録)が 100%収録されており、権利侵害関係調査においてはクレームや全文も 100%収録されているかどうかを最初に確認しておくことが重要です。

新興国の特許情報のようにデータが欠落した状態を認知してそれなりに調査せざるを得ない場合もありますが、その場合も欠落が収録のタイムラグであるのかも調査前に確認が必要です。

また、出願日基準で調査する場合には、直近のデータはデータ更新によって情報が追加されることを依頼者に知らせることも必須事項です。

以上