

2023年1月20日

台湾特許庁英語データベース収録検証

アジア特許情報研究会:伊藤徹男

1. はじめに

前報で台湾特許庁英語データベース (TPSS 英語 DB) が大きく変貌し、中国語の読み書きができない方でも英語インターフェースを元に調査可能になったことを紹介しました。

何しろこれまでの TWPAT 英語 DB での用語検索では「発明の名称」と「要約」から英語で検索し、表示も要約までしかなかったものが「請求の範囲」「詳細説明」まで検索、表示できるようになりましたから「大きく変貌」と表現しました。

そこでこの TPSS 英語 DB の英語情報の収録状況を早速確認し、実力の程度をお知らせしたいと思いました。

結論から先に申し上げますと、「がっかりです」。英語 DB というからには、また、「請求の範囲」「詳細説明」まで検索、表示できるようになりましたので、少なくとも商用英語データベース並みの英語情報が収録されたのでは、と期待しましたが、「TPSS 原語 DB に収録されているものとまったく同一の英語収録」ということが確認できました。

とはいえ、インターフェースが英語に変わりましたので初心者にとっては多少、とっつきやすいデータベースであり、ハードルが下がったものと思います。以下、収録データと共に紹介します。

2. TPSS 英語 DB の収録内容

まず、TPSS 原語 DB の英語収録状況を 2010 年～2022 年発行年で公開公報と、公告公報別に表 1 に示しました。出願年推移では、データが更新されるたびに収録内容が変化するために公報発行年推移としました。台湾では、公開は毎月 1 日、16 日に更新され、公告は 1 日、11 日、21 日に更新されます。したがって、発行年基準では 2022 年のデータは、2023 年 1 月時点ですべて発行され、データ更新によって追加されることはありません。

表 1. TPSS 原語データベースの英語収録状況

発行年	書誌 A	TI EN	収録率	AB EN	収録率	CL EN	収録率	発行年	書誌 B	TI EN	収録率	AB EN	収録率	CL EN	収録率
2010	44,962	22,265	50%	44,806	100%	7,044	16%	2010	16,345	5,802	35%	2,406	15%	2,334	14%
2011	46,157	22,680	49%	46,026	100%	6,759	15%	2011	20,025	7,283	36%	408	2%	2,696	13%
2012	51,592	24,988	48%	51,394	100%	7,040	14%	2012	25,535	10,573	41%	551	2%	3,837	15%
2013	52,126	25,827	50%	51,970	100%	7,251	14%	2013	40,251	16,998	42%	768	2%	6,219	15%
2014	48,719	23,948	49%	48,566	100%	6,941	14%	2014	45,602	19,398	43%	759	2%	7,442	16%
2015	47,367	23,775	50%	47,186	100%	6,684	14%	2015	48,315	21,127	44%	766	2%	7,684	16%
2016	44,356	22,274	50%	44,191	100%	6,991	16%	2016	48,947	20,882	43%	637	1%	7,283	15%
2017	43,677	22,026	50%	43,494	100%	6,563	15%	2017	45,711	18,850	41%	533	1%	6,751	15%
2018	44,071	22,624	51%	43,887	100%	6,849	16%	2018	36,147	14,625	40%	397	1%	5,801	16%
2019	47,987	24,708	51%	47,779	100%	6,924	14%	2019	34,928	13,676	39%	362	1%	5,429	16%
2020	46,829	28,119	60%	46,593	99%	6,945	15%	2020	33,813	15,924	47%	462	1%	5,370	16%
2021	47,938	29,295	61%	47,717	100%	7,575	16%	2021	36,578	17,464	48%	458	1%	5,843	16%
2022	49,553	28,469	57%	43,576	88%	7,355	15%	2022	37,175	18,323	49%	443	1%	6,007	16%

発明の名称(TI)、要約(AB)、請求の範囲(CL)別に左が公開特許、右が公告特許です。
 繁体字原語の収録は100%なので割愛しました。

公開特許の要約の英語収録は良好のようです。発明の名称の英語収録は50%程度ですが
 請求の範囲にはほとんどといっていいほど収録されていません。

TPSS 英語 DB の請求の範囲に英語情報が収録されていれば問題ないのですが、結果は表 2
 の如くです。全く同一です。

表 2. TPSS 英語データベースの英語収録状況

発行年	書誌 A	TI EN	収録率	AB EN	収録率	CL EN	収録率	発行年	書誌 B	TI EN	収録率	AB EN	収録率	CL EN	収録率
2010	44,962	22,265	50%	44,806	100%	7,044	16%	2010	16,345	5,802	35%	2,406	15%	2,334	14%
2011	46,157	22,680	49%	46,026	100%	6,759	15%	2011	20,025	7,283	36%	408	2%	2,696	13%
2012	51,592	24,988	48%	51,394	100%	7,040	14%	2012	25,535	10,573	41%	551	2%	3,837	15%
2013	52,126	25,827	50%	51,970	100%	7,251	14%	2013	40,251	16,998	42%	768	2%	6,219	15%
2014	48,719	23,948	49%	48,566	100%	6,941	14%	2014	45,602	19,398	43%	759	2%	7,442	16%
2015	47,367	23,775	50%	47,186	100%	6,684	14%	2015	48,315	21,127	44%	766	2%	7,684	16%
2016	44,356	22,274	50%	44,191	100%	6,991	16%	2016	48,947	20,882	43%	637	1%	7,283	15%
2017	43,677	22,026	50%	43,494	100%	6,563	15%	2017	45,711	18,850	41%	533	1%	6,751	15%
2018	44,071	22,624	51%	43,887	100%	6,849	16%	2018	36,147	14,625	40%	397	1%	5,801	16%
2019	47,987	24,708	51%	47,779	100%	6,924	14%	2019	34,928	13,676	39%	362	1%	5,429	16%
2020	46,829	28,119	60%	46,593	99%	6,945	15%	2020	33,813	15,924	47%	462	1%	5,370	16%
2021	47,938	29,295	61%	47,717	100%	7,575	16%	2021	36,578	17,464	48%	458	1%	5,843	16%
2022	49,553	28,469	57%	43,576	88%	7,355	15%	2022	37,175	18,323	49%	443	1%	6,007	16%

参考情報として、GPSS および商用データベースとして WIPS GLOBAL の台湾特許英語収録
 率をそれぞれ表 3 および表 4 に示しました。

表 3. GPSS の英語収録状況

発行年	書誌 A	TI EN	収録率	AB EN	収録率	CL EN	収録率	発行年	書誌 B	TI EN	収録率	AB EN	収録率	CL EN	収録率
2010	44,962	37,687	84%	44,956	100%	8,262	18%	2010	16,345	10,029	61%	2,621	16%	2,861	18%
2011	46,157	38,529	83%	46,151	100%	7,992	17%	2011	20,025	12,758	64%	687	3%	3,312	17%
2012	51,592	43,118	84%	51,589	100%	8,224	16%	2012	25,535	16,698	65%	837	3%	4,584	18%
2013	52,126	43,833	84%	52,119	100%	8,529	16%	2013	40,251	26,776	67%	1,201	3%	7,326	18%
2014	48,719	41,065	84%	48,710	100%	8,042	17%	2014	45,602	30,631	67%	1,172	3%	6,682	19%
2015	47,367	39,799	84%	47,318	100%	7,888	17%	2015	48,315	32,681	68%	1,144	2%	8,993	19%
2016	44,356	37,184	84%	44,328	100%	8,000	18%	2016	48,947	32,256	66%	1,020	2%	8,589	18%
2017	43,677	36,618	84%	43,634	100%	7,546	17%	2017	45,711	28,928	63%	842	2%	7,872	17%
2018	44,071	37,400	85%	44,038	100%	7,756	18%	2018	36,147	22,230	61%	637	2%	6,675	18%
2019	47,987	40,585	85%	47,918	100%	7,903	16%	2019	34,928	20,701	59%	598	2%	6,255	18%
2020	46,829	39,497	84%	46,690	100%	7,600	16%	2020	33,812	20,278	60%	565	2%	5,734	17%
2021	47,938	40,818	85%	47,793	100%	8,272	17%	2021	36,578	22,309	61%	560	2%	6,266	17%
2022	47,720	38,223	80%	42,511	89%	7,318	15%	2022	37,175	23,334	63%	504	1%	6,101	16%

GPSS の公開特許および公告特許の要約と請求の範囲の収録は TPSS と同等ですが、発明の
 名称では TPSS よりいくらか良好なようです。

表4. WIPS GLOBAL の英語収録状況

発行年	書誌 A	AB EN	収録率	CL EN	収録率	発行年	書誌 B	AB EN	収録率	CL EN	収録率
2010	44,962	44,916	100%	44,927	100%	2010	16,345	16,334	100%	16,271	100%
2011	46,157	46,106	100%	46,130	100%	2011	20,025	20,003	100%	19,986	100%
2012	51,592	51,543	100%	51,570	100%	2012	25,535	25,490	100%	25,502	100%
2013	52,126	52,085	100%	49,817	96%	2013	40,251	40,211	100%	37,884	94%
2014	48,720	48,701	100%	47,800	98%	2014	45,602	45,544	100%	45,585	100%
2015	47,367	47,365	100%	47,275	100%	2015	48,315	48,278	100%	48,307	100%
2016	44,356	44,352	100%	44,272	100%	2016	48,947	48,928	100%	48,806	100%
2017	43,677	43,670	100%	43,659	100%	2017	45,711	45,694	100%	45,578	100%
2018	44,071	44,069	100%	44,063	100%	2018	36,147	36,146	100%	33,999	94%
2019	47,987	47,983	100%	47,970	100%	2019	34,928	34,919	100%	34,914	100%
2020	46,829	46,826	100%	46,800	100%	2020	33,812	33,810	100%	33,798	100%
2021	47,939	47,939	100%	47,916	100%	2021	36,578	36,573	100%	36,555	100%
2022	47,712	47,705	100%	38,777	81%	2022	37,175	36,041	97%	29,613	80%

WIPS GLOBAL は商用データベースだけあって要約、請求の範囲の収録も良好のようです。英語で検索する場合には、やはり商用英語データベース、ということになりますが、英語も繁体字（中国語）も異表記を網羅して漏れのない検索をするには把握している限りの「英語異表記+繁体字異表記」のハイブリッド検索がお勧めです。

3. データベースの収録タイムラグ

次いで、原語（繁体字）情報の収録タイムラグはどの程度かについても見てみました。

表5. 各データベースの原語収録タイムラグ

発行年	TPSS		GPSS		WIPS GLOBAL	
	A	B	A	B	A	B
2000		15,656		15,656		15668
2001		24,429		24,429		24438
2002		23,036		23,036		23037
2003	8,194	25,134	8194	25,134	8194	25141
2004	28,927	20,457	28927	20,457	28927	20461
2005	41,439	20,632	41439	20,632	41438	20630
2006	44,783	23,229	44783	23,229	44783	23228
2007	46,986	22,219	46986	22,219	46986	22218
2008	50,141	12,867	50141	12,867	50141	12867
2009	52,618	14,139	52618	14,139	52618	14141
2010	44,962	16,345	44962	16,345	44962	16345
2011	46,157	20,025	46157	20,025	46157	20026
2012	51,592	25,535	51592	25,535	51592	25537
2013	52,126	40,251	52126	40,251	52126	40254
2014	48,719	45,602	48719	45,602	48720	45604
2015	47,367	48,315	47367	48,315	47367	48322
2016	44,356	48,947	44356	48,947	44356	48950
2017	43,677	45,711	43677	45,711	43677	45713
2018	44,071	36,147	44071	36,147	44071	36150
2019	47,987	34,928	47987	34,928	47987	34927
2020	46,829	33,813	46829	33,812	46829	33812
2021	47,938	36,578	47938	36,578	47939	36577
2022	49,553	37,175	47720	34934	47712	36043

公開特許は2003年から収録が始まりましたが、公告特許については2000年以降の収録を確認しました。公告特許においては2008年～2010年にはその前後の年度より少ない収録が確認できましたが、審査制度など法律的な観点で何か変化があったかもしれない（未考察です）。

やはり GPSS および商用データベースにおいて2022年に収録タイムラグと見られる収録不足が確認できたので、さらに2022年について1か月ごとに、また12月については公開、公告とも発行日ごとの収録を確認しました。（データ取得は2023年1月3日です。）

表6. 原語収録タイムラグ

発行日	TPSS原語		GPSS原語		WIPS GLOBAL原語	
	A	B	A	B	A	B
202201	3,742	2,748	3,742	2,748	3,741	2,748
202202	4,046	2,317	4,046	2,317	4,045	2,317
202203	3,986	3,094	3,986	3,094	3,986	3,094
202204	4,152	3,177	4,152	3,177	4,150	3,177
202205	4,603	3,414	4,603	3,414	4,603	3,414
202206	4,006	3,276	4,006	3,276	4,006	3,277
202207	3,953	3,125	3,953	3,125	3,953	3,125
202208	4,539	3,471	4,539	3,471	4,539	3,471
202209	3,889	3,213	3,889	3,213	3,888	3,212
202210	4,305	2,969	4,305	2,969	4,305	2,966
202211	4,333	3,018	4,333	3,018	4,331	3,018
202212	3,999	3,353	2,166	1,112	2,165	2,224
20221201	2,166	1,112	2,166	1,112	2,165	1,112
20221211		1,112		0		1,112
20221216	1,833		0		0	
20221221		1,129		0		0

GPSS では12月16日発行の公開特許、12月11日および21日発行の公告特許が未収録であり、WIPS GLOBAL では12月16日発行の公開特許、12月21日発行の公告特許が未収録でした。WIPS GLOBAL と同様の収録タイムラグは他の商用データベースでも見られました。

2節の英語収録状況、3節の収録タイムラグでは、既に紹介しました「データベース収録確認ワード（その1）」¹⁾、「データベース収録確認ワード（その2）」²⁾で検証しました以下の用語で実施しました。

AB原語：「。 or 一 or 的 or 種 or 以 or 之 or 本」

CL原語：「1 or 。 or 一 or 種 or 其 or 以 or : or 、 or 之」

英語：「acc* or aci* or act* or ada* or add* or adj* or aro* or arr* or acc* or ass* or can* or com* or con* or cos* or def* or det* or dis* or eac* or ele* or ena* or enc* or end* or exa* or ext* or fib* or fil* or fir* or hav* or inc* or inv* or loa*

or loc* or met* or mor* or pho* or pla* or plu* or pol* or por* or pre* or pro* or rea* or rec* or rep* or ret* or sea* or sec* or sem* or sep* or ser* or str* or sub* or sup* or sur* or ther* or tra* or tre* or use*]

4. 具体例での検証比較

最後に、具体的な用語、IPC 分類、出願人名での各データベースの収録状況を確認しました。ここでも TPSS 英語 DB の収録は TPSS 原語 DB と同一と考えられますので TPSS 原語 DB で検証したデータを表 7 に示しました。

いずれも網羅的ではありませんが、3D PRINTER の英用語①、繁体字②、代表的 IPC④および代表的出願人⑤を、用語は検証が容易なように発明の名称のみで 2010 年～2022 年発行の公開特許を抽出しました。

①(3D PRINT* or THREE DIMENSION* PRINT*)

②(3D 列印 or 三維列印 or 3D 印表 or 三維印表 or 3D 印刷 or 立體列印)

③ ① or ②

④(B29C064* or B29C067* or B33Y*)

⑤三緯國際立體列印科技 (XYZPRINTING)

表 7. 具体例での比較

	TPSS 原語DB		GPSS		WIPS GLOBAL	
	原語	英語	原語	英語	原語	英語
①		465		465		500
②	441		441		437	
③	488		488		524	
④	1328		1328		1325	
⑤	140	140	140	140	140	140

その結果、TPSS と GPSS では同一の結果となりました（両方とも台湾特許庁データベースですから当然ですが）。

英用語では WIPS GLOBAL > TPSS, GPSS、繁体字では TPSS, GPSS > WIPS GLOBAL となりました。それぞれ差分について検証しましたが、有意差を確認できなかった。

尚、WIPS GLOBAL における英語検索では、“3D PRINT*”、“THREE DIMENSION*”のようにフレーズはダブルクォーテーションで囲って、ハイフンの有無で検索結果が異なるので「“THREE DIMENSION*” or “THREE-DIMENSION*”」のようにハイフンを含むものも OR 演算しました。

5. まとめ

台湾特許庁英語データベースには原語情報、英語情報とも原語データベースと同じものが収録されましたが、英語データベースだからと英語収録が充実された訳ではないことも確認できました。今後、英語収録が充実されることを期待したいと思います。

権利判断調査や無効化資料調査のような網羅的な調査でなく、出願前の先行技術調査など研究者・技術者が気軽にいつでもどこでもアクセスできる無料データベースとしては有用なツールと思います。

ある程度の網羅性も考慮すると現時点では原語情報を収録する商用データベースで「英語＋原語」で検索することが賢明だと思います。但し、直近1か月程度は収録タイムラグもあるのでSDIなどではその点も考慮して情報収集する必要があります。

参考文献

1) データベース収録確認ワード(その1)

http://patentsearch.punyu.jp/asia/CoverageWord_EN.pdf

2) データベース収録確認ワード(その2)

http://patentsearch.punyu.jp/asia/CoverageWord_CNTW.pdf

以上