

台湾・中国特許情報の文字化け文字

アジア特許情報研究会：伊藤徹男

1. はじめに

データベース中の中国語（繁体字、簡体字）が「季・」(quaternary phosphonium)、
「∴化合物」(phosphonium compounds)のように正常に表記されない文字を「文字化け文字」として紹介する。

文字化け文字は台湾特許の化合物において特に多く出現するので、まず台湾特許について検証した。検証は、台湾特許、中国特許共に台湾特許庁のワールドワイドな特許情報を収録する GPSS および中国特許の一部は WIPO の PATENTSCOPE を利用した。

GPSS の英語訳付与が不十分な点もあるので十分な検証は難しいが、後日、英語翻訳が充実している商用データベースで再検証することとしたい。

台湾特許における「文字化け文字」については、2015 年にアジア特許情報研究会と台湾特許庁との情報交換においてデータベースの改善項目として要望したもののうちの 1 つである。ほとんどの要望は即座に聞き入れていただいたが、ここで取り上げる文字化け文字については台湾で使用する PC では問題なく表記されるようであり、特殊な文字コードで「日本で使っている PC 環境では表示できない」ということのようなのである。

また、いずれも 2009 年 9 月以前に出願された古いものでもあり、台湾特許庁としてはこれら文字化け文字が日本の PC 環境でも閲覧できるようにという要望には応えられない、との回答をいただいた。台湾特許データベースへの英語付与は不十分ながらほぼ問題なく英訳もされている点を考慮すると台湾では文字化けすることなく表記されていることが窺える。

したがって、ここで紹介する文字化け文字は英訳が充実している商用英語データベースを使えば議論するに値しないものと思われるが、繁体字や簡体字（原語）と英語を組み合わせたハイブリッド検索において原語の異表記として加える際に、このような文字化け文字もある、という程度の認識でよいかもしれない。

以下の検証ではその存在数を公開特許数と公告特許数で次のように示した。

TI:23/101（発明の名称中の「公開特許 23 件/公告特許 101 件」）

検証は 2022 年 12 月末に実施したが、文字化け文字の多くは 2009 年 9 月出願までであり、中には 2000 年以前出願のものが多い。特に 2002 年出願以前のものには権利侵害調査の観点からは対象外とも言えるが、参考情報として紹介しておくこととしたい。

2. 台湾特許における文字化け文字

1) indol 検証

INDOLE は繁体字(簡体字も同じ)で「吲哚」と表記される。

吲哚化合物(INDOLE COMPOUNDS), 吲哚類衍生物(INDOLE DERIVATIVES)

「基」が付加した「吲哚基」は INDOLYL と英訳され、DIAZAIN~~INDOLYL~~PYRIDONE(二氮雜~~吲哚基~~吡啶酮)のように英語では中間一致検索が必要となることも多い(中国語では部分一致検索となるので、前方一致、中間一致、後方一致検索などの意識は不要)。

また、吲哚酮(INDOLONE), 1H-吲哚-5(1H-INDOL-5), 吲哚啉(INDOLINE)などと英訳もされるので、英語では「*INDOL*」と中間一致検索する必要がある。

「*INDOL*」と検索したものの中には、その存在数は極めてわずかではあるが、以下のような文字化け文字が確認できる。

a) 吲・(INDOLE) TI:23/101, AB:4/9, CL:0/10 2002 年出願以前

異吲・酮(ISOINDOLONE), 吲・啉(INDOLINE), 吲・基(INDOLYL)

上記「・」(この場合には「哚」が化けたもの)を含んでコピペして検索すると、ちゃんと「異吲・酮」などを抽出できる。

「*INDOL*」で検索すると吲嗪(INDOLIZINE)が抽出されると同時に、「嗪」が化けた「吲・」「吲・」「吲・」の文字化け文字も抽出される。ワード上では「吲・」で表される文字化け文字も INDOLE で確認された「吲・」とは異なるのでやっかいである。さらに、中国語異表記として嗪が欠落した吲酮(INDOLIZINONE)などもある。

GPSS の検索画面でもワード画面と同様、図 1 のように区別がつかないが、エクセルや文字コードを UTF-8 形式で保存したテキスト画面では表 1 のように哚や嗪が文字化けした 4 種の文字化け種が区別されている。通常の ANSI 形式で保存したものは文字コード情報が消えてしまうので文字化け文字を保持するには文字コードを意識して保存する必要がある。

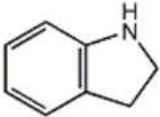
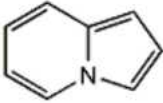
(吲□ or 吲□ or 吲□ or 吲□)@ti

図 1. INDOLE, INDOLIZINE 4 種の文字化け文字

表 1 のエクセル画面の a) と d) はやはり同一に見えるが、OR 検索してみると重複していないことがわかる。

a) 吲・ or d) 吲□ (INDOLE or INDOLIZINE) TI:23/103

表 1. 吲哚(INDOLE), 吲嗪(INDOLIZINE)の存在数

	TI	AB	CL	
	665/342	810/766	5151/2804	
a) 吲・	23/101	4/9	0/10	indole(吲哚)
	7/4	3/3	40/23	
b) 吲	2/0	0	0	
c) 吲	2/2	0/2	2/1	
d) 吲・	0/2	1/6	0	indolizine(吲嗪)

さらに、吲哚や吲嗪の2文字共に化けてしまった表2のようなものも存在する。やはり、その存在数は極めてわずかでクレームだけに存在するようです。

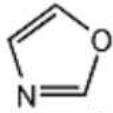
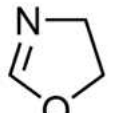
表 2. 吲哚および吲嗪の2文字共化けている例

	TI	AB	CL
e) 𐄂	0	0	77/24
f) 𐄃	0	0	3/9
g) 𐄄	0	0	2/3
h) 𐄅	0	0	14/70

2)oxazol 検証

「*OXAZOLE*」で検索すると文字化け文字を含む「噁唑, 唑, ・唑, ・唑, ・唑」を抽出できるが、異唑 (ISOOXAZOLE), 唑化合物 (OXAZOLE COMPOUND) など、「唑」一文字で OXAZOLE と英訳されているものもある。

表 3. OXAZOLE, OXAZOLINE の文字化け文字

	TI	AB	CL	
a) 噁唑	96/211	393/363	4140/2388	
b) 𐄆	197/157	244/349	482/103	oxazol
c) ・唑	4/11	1/39	107/103	
d) ・唑	1/0	0	14/67	
e) ・唑	0	0	30/137	oxazoline

a)には、異噁唑 (ISOOXAZOLE), 噁唑啉 (OXAZOLINE), 噁唑酮 (OXAZOLONE), 噁唑啉酮 (OXAZOLIDINONE), 苯並噁唑, 苯并噁唑 (BENZOXAZOLE) などを含む。

b)には、異噁唑(ISOXAZOLE)、異噁唑啉(ISOXAZOLINE)などを含む。

c)には、異・噁(ISOXAZOLE)、異・噁啉(ISOXAZOLINE)などを含む。

OXAZOLINE(噁唑啉)では噁を含まない唑啉、また、唑啉が OXAZOLINE として、異唑啉、異唑が ISOXAZOLINE として抽出される。

OXAZOLINE が「唑烷」と表記されるものとして、唑烷(OXAZOLIDIN)、噁唑烷(OXAZOLIDINE)や酮(KETONE)が付加した唑烷酮(OXAZOLIDONE)、噁唑烷酮(OXAZOLIDINONE)も多い。


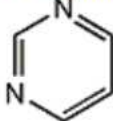
3) pyridine 検証

吡啉(PYRIDINE)の中国語には誤字と思われるような異表記がいくつか存在するが、文字化け文字としては「X啉」があり、1998年以前の公告公報には多数存在する(表4)。

「X啉」はX啉酮(PYRIDONE)、X啉基(PYRIDYL)と英訳されることが多く、*PYRIDIN*では抽出されず、*PYRID*と少し短く中間一致検索しないとカバーできない。

また、嘍啉(PYRIMIDINE)が PYRIDINE と誤訳されているケースもいくつか存在する。

表4. PYRIDINE の異表記と文字化け文字

	TI	AB	CL	
a) 吡啉	1835/1177	2533/2743	16701/11391	 pyridine
b) 砒啉	3/1	8/6	64/46	
c) 比啉	1/1	2/1	2/1	
d) 比啉	7/1	8/2	64/52	
f) 嘍啉	0/25	0	0	 pyrimidine
g) 嘍啉		0	0	
X啉	0/747	0	0	

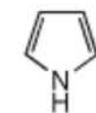
文字化け文字「X」について調べたのが表5である。

最も多いのは PYRIDINE(*PYRID*)であるが、「X」のすべては 1998 年以前の出願であり、権利侵害調査の観点からは無視できるが、それにしても文字化け文字としては多いように思う。PIPERIDINE と PIPERAZINE については次節で紹介する。

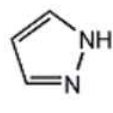
表 5. 文字化け文字 X

	TI	AB	CL	
h) X	0/1474	4/18	16/40	
i) X 啖	0/747	0	0	pyridine(吡啶)
j) X 咯	0/285	0	0	pyrrole(吡咯)
k) X 唑	0/223	0	0	pyrazole(吡唑)
l) X 喃	0/109	0	0	pyran(吡喃)
m) 六氫 X 啖	0/58	0	0	piperidine(哌啶)
n) 六氫 X	0/43	0	0	piperazine(哌嗪)

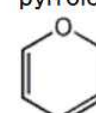
すべて1998年以前の出願



pyrrole



pyrazole



pyran

4) piperidine と piperazine 検証

PIPERIDINE と PIPERAZINE の異表記と文字化け文字を表 6 および表 7 に示した。

表 6. PIPERIDINE の異表記と文字化け文字


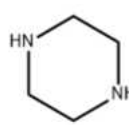
	TI	AB	CL	
a) 哌啶	187/118	524/517	5614/2967	 piperidine
b) 哞啶	7/3	19/6	247/179	
c) ▶ 啶	76/64	19/14	62/17	
d) ρ啶	1/0	0	0	
e) ・啶	0/100	0	0	
f) 六氫吡啶	201/171	218/406	2066/1613	
g) 六氫吡啶	0/1	1/1	0/8	
h) 六氫 X 啶	0/58	0	0	

表 7. PIPERAZINE の異表記と文字化け文字

	TI	AB	CL	
a) 哌嗪	54/42	182/194	2243/1165	 piperazine
b) 哌𠄎	37/50	24/74	12/10	
c) 哌・	4/13	58/43	0/27	
d) ▶ 𠄎	57/9	4/1	45/2	
e) ▶ 𠄎	21/14	6/4	0/2	
f) ▶)	0/3	0/1	0/2	
g) 六氫吡𠄎	59/73	19/45	24/93	<p>六氫X啶をNOT</p>
h) 六氫吡・	9/20	48/157	1/48	
i) 六氫 X	0/43	0	0	

GPSS では、英語付与が不十分であるため、正確な数字は把握できないが

PIPERIDINE として使われている中国語が PIPERAZINE と英訳されることはないか (その逆も) など、およその誤訳状況を調べ、表 8 に示した。

表 8. PIPERIDINE と PIPERAZINE の誤訳 (発明の名称のみ)

	piperidin	*piperazin*
哌啶	156/88	1/0
呱啶	7/0	0
▶ 啶	57/50	0
六氢吡啶	170/130	0
哌·	1/0	2/12
哌 □	1/0	27/39
六氢吡 □	0	36/48
六氢吡·	0	4/14

表 8 の値は以下のように求めた。

((哌啶 and *piperidin*) not *PIPERAZIN*)@ti TI:156/88

((哌啶 and *PIPERAZIN*) not *piperid*)@ti TI:1/0

相互に誤訳があるように見えたが、表 8 を見る限りは大きな誤訳はないようである。

表 8 では、PIPERIDINE と PIPERAZINE の総計が中国語と差異があるのは英語付与が不十分な点に加え、以下のように両方が併記されているものは除いたためである。

- ・ 哌嗪-哌啶(PIPERAZINE-PIPERIDINE)
- ・ 哌嗪(啶)(PIPERAZINE (PIPERIDINE))
- ・ 哌啶及哌· (PIPERIDINE AND PIPERAZINE)
- ・ · · · 及 · 啶(PIPERAZINE & PIPERIDINE)
- · · · · 啶(PIPERAZINE-PIPERIDINE)

権利侵害調査は別として、類似構造の PIPERIDINE と PIPERAZINE は中国語および英語とも OR 検索でほぼ問題はない (極端にノイズが増えるものではない)。

5) その他の文字化け文字

文字化け文字は、上記したように主に N 原子を含む複素環化合物に多いが、O 原子を含む複素環化合物やアセトニトリルなど非環状のものにもある。

a) furan など O 含有化合物

多くは発明の名称や要約中ではなくクレーム中に以下の文字化け文字が存在する。

呋喃(FURAN)では、□喃，・喃，□喃などの文字化け文字を確認でき、四氫呋喃(TETRAHYDROFURAN)では、四氫 喃(呋がスペースになって欠損)、二氫喃(呋が欠損)したものその他、四氫口夫喃(呋が口夫となった誤字)などもある。

BENZOFURAN では苯并呋喃，苯并呋喃，苯骈呋喃，氧芴などの異表記の他、クレーム中に苯・の文字化け、発明の名称には「苯並 喃」など空白があるものを確認できる。

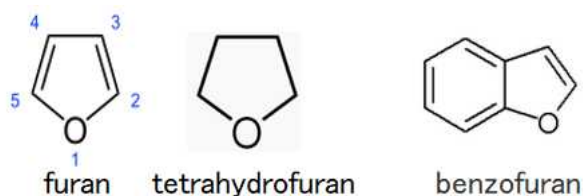


図 2. FURAN

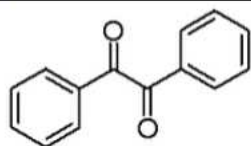
醌(QUINONE)では、苯□(BENZOQUINONE)，□□ (NAPHTHOQUINONE)などクレーム中に 2 種の文字化け文字「□, □」を確認できる。

b) benzyl

BENZYL(苯甲基，甲苯基，苄基，苄基，苄甲基，苄甲基)では表 9 のような文字化け文字を確認できるが、・(BENZYL)のみではクレーム中に公開 160 件/公告 355 件が存在する。

表 9. BENZYL の文字化け文字

	TI	AB	CL	
苄基, 苄基				BENZYL
⊕基	0	0	119/254	
苄氧基				BENZYLOXY
⊕氧基	0	0	37/42	
苄胺, 苄基胺				BENZYLAMINE
⊕胺	0	0	5/12	
丙烯酸苄酯				BENZYL ACRYLATE
丙烯酸⊕酯	0	0	4/3	



c) nitrile

*aceto nitrile or *acetonitrile と検索すると、乙腈の他、「乙・ or 乙・」の文字化け文字や空白が入った「乙 的(acetonitrile)」が得られる。

胺基乙□(AMINOACETONITRILE), 苯基乙・(phenylacetonitrile)

それぞれ発明の名称では、乙□(公開 2 件/公告 1 件)、乙・(公開 1 件/公告 13 件)と検索され、2 種の□と・は重複はなく、nitrile と英訳されている。

最初の□(nitrile)では発明の名称で公開 19 件/公告 19 件で□橡膠(NITRILE RUBBER), 甲□(CARBONITRILE), 磷□酸(phosphonitrilic acid)などが抽出され、後の・(nitrile)では公開 132 件/公告 150 件で螺環・(SPIROCYCLIC NITRILE), ・橡膠(NITRILE RUBBER), 丙烯・(ACRYLONITRILE)などが抽出される。したがって 2 種の文字化け nitrile は繁体字で検索する場合には nitrile の異表記として OR 検索する必要がある。

2 種の nitrile 文字化け文字も、やはり 2009 年出願以前のものである。

不飽和腈、丙烯腈、芳族腈などのニトリル化合物も不飽和 (unsaturated nitrile)、丙烯 (acrylonitrile)、芳族 (aromatic nitriles)のように nitrile 部が空白であるが nitrile と英訳されている。

3. 中国大陸特許における文字化け文字

これまでは台湾特許の文字化け文字を紹介してきたが、中国大陸特許の文字化け文字は台湾特許ほどではないが存在する。

その数例を表 10 および図 3、図 4 に示す。

表 10. 中国大陸特許の文字化け例

	TI	AB	CL	
・	0	0	150/56	磷・(phosponium)
・	0	16/25	6/20	磷・(phosponium)
∴	647/225	221/174	221/174	∴唑(*oxazole)
・	30/ 7	29/ 11	20/13	・基(onium group)
★	3/1	4/0	4/4	有机★盐(organophosponium)
□ ↑	3/2	0	1/0	□金属(onium metal)
*				磺酸*盐(onium sulfonate)

文字化け文字「*」についての存在数は空欄であるが、GPSS における検索では「*」はワイルドカードとして働くために確認できていない。

図 3 と図 4 に phosphonium の文字化け例を示す。

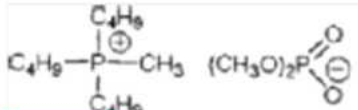
CN200480044203.5

季 \square 氢氧化物: CNIPRでは未翻訳 \Rightarrow PATENTSCOPEでは季 \square 磷氢氧化物
J-PlatPatでは「第4級の・の水酸化物」

CN200680021175.4

磷 CNIPRでは未表示 \Rightarrow PATENTSCOPE要約にイメージあり
J-PlatPatでは「スルホン酸の・の塩」
硫酸 \square 盐: phosphonium sulfonate

CN200710111843.9



季 \square 盐: quaternary salt

CNIPRでは未表示
 \Rightarrow PATENTSCOPE要約にイメージあり
J-PlatPatでは「第4級の・の塩」

CN200880004968.4

J-PlatPatでは「4ヒドロキシメチルの★の化合物」

$P^{\oplus}(CH_2OH)_4 Cl^{\ominus}$ CNIPRでは未表示 \Rightarrow PATENTSCOPEにイメージあり

四羟甲基 \square 化合物: tetrakis(hydroxymethyl)phosphonium compound

ほとんど要約中に存在 \square : 要約中に16件

29

図3. 中国特許情報中の文字化け文字(phosphonium)

上記、中国語文字化け4件のいずれも J-PlatPat では日本語に翻訳されていない。

1. 环氧-有机硅混合树脂组合物及发光半导体装置 发明专利 有效

申请号: CN200710111843.9 申请日: 2007.06.15
 公开(公告)号: CN101089048A 公开(公告)日: 2007.12.19
 同日申请: 分案原申请号:
 申请(专利权)人: 信越化学工业株式会社
 分类号: C08L83/04(2006.01); C08L83/00(2006.01); C08K5/53(2006.01); C09K3/10
 优先权: 2006.06.16 JP 2006-167072

摘要: 环氧-有机硅混合树脂组合物, 其特征在于, 以(A)-分子中具有至少2个环氧官能性基团或氧杂环丁烷基的有机硅烷半氧烷树脂, (B)-分子中具有至少2个环氧官能性基团的环氧树脂, (C)固化剂(D)固化催化剂作为必须成分, 固化催化剂含有1种以上 \square 季盐。本发明的环氧-有机硅混合树脂组合物, 通过用其固化物被覆保护发光半导体元件, 由于再流试验中无变色, 故可以提供安装可靠性优异的发光半导体装置, 产业上的优点很多。这种情况下, 玻璃化转变温度在130°C以上的固化物, 特别是在固化物表面完全不附着尘埃, 在耐热试验中耐开裂性优异。

本篇专利无摘要附图

CNIPR要約

EN Epoxy-organic silicon mixed resin combination and luminous semiconductor device
ZH 环氧-有机硅混合树脂组合物及发光半导体装置

Abstract: **EN** The epoxy organosilicon mixing resin composition is characterized in that it includes (A) an organic polysiloxanes which includes at least two epoxy functional groups in one molecule, (B) an epoxy resin which includes at least two epoxy functional groups in one molecule, (C) a solidifying agent, and (D) a solidifying catalyst as the necessary components, and the solidifying catalyst contains at least one \square quaternary salt. The epoxy organosilicon mixing resin composition of the present invention protects the semiconductor component by the condensate, it can provides the semiconductor device with excellent reliability as there is no color changing in the second scotching experiment and there are many advantages in the industry. In this situation, the glass transition temperature is above 130 centigrade will not ad surface, and the cracking-resistant property is e
ZH 环氧-有机硅混合树脂组合物, 其特征在于, 以(A)-分子中具有至少2个环氧官能性基团或氧杂环丁烷基的有机硅烷半氧烷树脂, (B)-分子中具有至少2个环氧官能性基团的环氧树脂, (C)固化剂和(D)固化催化剂作为必须成分, 固化催化剂含有1种以上 \square 季盐。本发明的环氧-有机硅混合树脂组合物, 通过用其固化物被覆保护发光半导体元件, 由于再流试验中无变色, 故可以提供安装可靠性优异的发光半导体装置, 产业上的优点很多。这种情况下, 玻璃化转变温度在130°C以上的固化物, 特别是在固化物表面完全不附着尘埃, 在耐热试验中耐开裂性优异。

PATENTSCOPE要約

(1) $C_6H_5-P^{\oplus}(C_6H_5)_2-CH_3$ $(CH_3O)_2P^{\ominus}SO_3$

(2) $C_6H_5-P^{\oplus}(C_6H_5)_2-CH_3$ $(C_6H_5O)_2P^{\ominus}SO_3$

図4. CNIPR と PATENTSCOPE の表記(AN:CN200710111843.9)

CNIPR では文字化けしていても PATENTSCOPE では正常に表記され、また正常に英訳もされている。

4. まとめ

以上見てきたように、いずれの文字化けも問題なく英訳されているようなので英語で検索する場合には問題はなさそうである。やはり台湾など現地の環境では問題なく表記されるのでその原語から正しい翻訳もなされているものと思う。

英語情報から洩れる「文字化け文字」も極めて少ないので、調査という観点からは無視してもよいものと思う。したがって、このような現象もある、という情報として受け止めていただければ結構です。

実務にも役に立たないそんなマイナーなことを、と冷笑されそうである。

但し、現時点でも中国大陸特許の英訳には誤訳、スペルミスなどは少なからず存在するため、検索では「原語（中国語）＋英語」が必須であることに変わりありません。

台湾特許や中国大陸特許の文字化け文字は多種にわたり、すべて検証して紹介できなかったことをお断りしておきます。

参考情報

中国大陸特許の原語表記には以下のようなものもある(CN200610010053.7)。

要約： 本发明涉及计算机与中文笔形，具体为一种中文笔形输入法。其特点是：中文笔形经计算机键盘的1、2、3、4、5、6、7、8、9、0键输入，具体对应是，横横1——、' 笔形，竖2——|、二、||、v、)、"、く、口、口笔形，撇3——ノ、口、彡、《《、彡、三、口、川笔形，点4——、"、"笔形，折5——口、|、フ、L、口笔形，叉6——义、六笔形，穿7——卅、廿、丰、口笔形，框8——井、口、八、口笔形，围9——口、口、長、口、口、口、口笔形，封0——口"、十笔形。中文笔形输入法，主要用于汉字的输

以上