

特許統計データベースの課題 —OECD特許統計タスクフォースの取り組み—

経済協力開発機構 岡崎 輝雄

1. はじめに

前号(特技懇No.249号p.104-113)で紹介しましたように、経済協力開発機構(OECD)では、加盟国代表で構成される科学技術政策委員会(CSTP; Committee for Scientific and Technological Policy)と産業イノベーション起業委員会(CIIE; Committee on Industry, Innovation and Entrepreneurship)で、各国政府の付託を受けて、イノベーション戦略プロジェクトを実施しています¹⁾。両委員会とも、イノベーションについての基礎的なデータとして二国間貿易統計(BTD; Bilateral Trade Database by industry)や産業連関表(I-O; harmonised Input-Output tables)などを蒐集したSTAN database (Structural Analysis database)²⁾に基づき、新たなイノベーション政策を描きつつありますが、こうした中で、特許統計は、特許と企業業績/生

産性のマイクロ分析、出願人や発明者の住所情報を活用したイノベーション集積/分布に関する空間経済分析、引用文献情報を用いた技術連関分析など、このプロジェクト成否の鍵を握っています。

OECDでは、上記STAN databaseと並行して、日本国特許庁、欧州特許庁、アメリカ合衆国特許商標庁、世界知的所有権機関、欧州委員会、全米科学財団からなる特許統計タスクフォース³⁾の協力のもと、特許統計データベース、PATSTAT (Worldwide Patent Statistical Database)の開発を進めてきました。

2. PATSTATとは

特許統計を研究開発成果やイノベーション活動の指標として利用するためには、精緻な特許統計データベースが必要です⁴⁾。特に国境を越える研究開発活動

1) 科学技術政策委員会(CSTP; Committee for Scientific and Technological Policy)では、その下部組織であるイノベーション・技術政策作業部会(TIP; Working Party on Innovation and Technology Policy)において特許政策を、科学技術指標専門家会合(NESTI; Working Party of National Experts on Science and Technology Indicators)において特許指標を、バイオテクノロジー作業部会(WPB; Working Party on Biotechnology)において遺伝子関連特許のライセンスに関するガイドラインを、産業イノベーション起業委員会(CIIE; Committee on Industry, Innovation and Entrepreneurship)では、模倣品関連やソフトウェア特許のあり方を議論している。いずれのテーマについても、特許統計は議論の裏付けに欠くことのできない要素であり、OECDの政策立案・意思決定のための調査と分析において、その重要性は急速に高まっている。

2) http://www.oecd.org/document/15/0,3343,fr_2649_201185_1895503_1_1_1_1,00.html

3) OECDは日本国特許庁、欧州特許庁、アメリカ合衆国特許商標庁、世界知的所有権機関、欧州委員会、全米科学財団からなる特許統計タスクフォースを設置している。PATSTATを体系的なデータベースとして構築することを目指しつつ、特許統計タスクフォースは2006年10月にオーストリア特許庁にて、2007年5月に世界知的所有権機関にて、2007年10月にベネチア国際大学にて国際会議を開催した。タスクフォースメンバーによる会議とともにカンファレンスも開催され、ユーザサイドの視点からPATSTATのデータベースを用いた特許統計等の報告が行政官、研究者よりなされている。

4) 和田哲夫、後藤晃、「特許に関する経済学的な分析の必要性和特許データのありかた」、産業界におけるライセンス契約の動向及び経済学的課題に関する調査研究報告書、財団法人知的財産研究所、p.148-159、平成15年3月

を特許統計から捕捉するためには、世界規模でのデータベースが整備されなければなりません。

2006年4月に初めてリリースされたPATSTATは、世界80箇国⁵⁾から蒐集された5千万以上もの出願情報を有しており、現在、欧州特許庁が窓口となって頒布しています。PATSTATはリレーショナルデータベース⁶⁾であり、各データがコンマで区切られた16のCSVファイルから構成され、欧州特許庁の書誌データベース (DocDB)の一部として、出願番号、公開番号、優先日や出願日、公開日、出願人や発明者の名前と住所、国コード、発明の名称、国際特許分類、ファミリー、要約、非特許文献を含む引用文献情報などが記録されています。

PATSTATは、データカタログとセットで毎年春と秋の2回、DVD2枚組で配布されていますが、その利用を学術に限っており、商用目的には使用できません。データベースの普及ポリシーや技術的な課題について、年1、2回のペースで特許統計タスクフォースにより議論されています。

3. PATSTATの抱える課題

OECDでは、新しいイノベーション政策に向けて、特許統計を過大にも過小にも評価しないよう、PATSTATが内包する誤差についての検討を進めています。PATSTATは、特許統計を扱うユーザにとって有用なデータベースとなりましたが、利用に当たっては留意点があります。

PATSTATのデータソースは各国特許庁で定められた言語と書式で構成された出願情報であり、それらは、特許統計を意図して用意されたものではありません。したがって、データ構造の最適化なしに、各国特許庁の出願情報をそのまま特許統計に利用したのでは、ユーザにとっては使い勝手が悪い、ミスリーディングをおこしかねないデータベースとなってしまいます。

PATSTATの抱える課題の一つに各国毎のデータ項目が完全には統一されていない点が挙げられます。

例えば、第1表に記載されるように、欧州特許庁やアメリカ合衆国特許商標庁の出願情報には、国コードが存在するのに対し、日本国特許庁の出願情報には国コードが含まれません。PATSTATに収められている日本国特許庁の出願情報には出願人、発明者の英文住所データが十分に含まれておらず、このことは、国際比較を行う場合やイノベーション集積分布に関する空間経済分析を行うような場合において支障ともなっています。

PATSTATにはこの他にも、特許統計上、留意しなければならない点が存在します。その一つが出願人名の表記ゆれです。第1表はPATSTATから無作為に抽出し、対応するファミリーを並べたものです。同じファミリーであっても、三極特許庁間で出願人名の表記に違いがあることが分かります。

4. 出願人名の表記ゆれ

一つの出願人名に対して、微妙な表記の差異があったり、複数通りの表記が存在したり、同じファミリーに属しながら異なる出願人名が記載されていたりと、PATSTATには大なり小なりデータ上の不整合が存在します。原因として次のような理由が考えられます。

(1) 表音表記

第1表には、SONII、SHERU、EREKUTORON、BURIJISUTONといった出願人名が記載されていますが、これらは、ソニー、シェル、エレクトロン、ブリヂストンを表音表記したものと考えられます。ソニー、シェル、エレクトロンはそれぞれSONY、SHELL、ELECTRONと表記するのが一般的でしょうし、単純なローマ字表記であれば、ブリヂストンはBURIDISUTONとつづるべきかもしれません。

一般に、アルファベットでない言語をアルファベットに置き換える際、長音を加えたり、故意に英語発音に模したつづりにすることで、表記ゆれが頻出します。

5) [http://documents.epo.org/projects/babylon/eponet.nsf/0/2464E1CD907399E0C12572D50031B5DD/\\$File/global_patent_data_coverage.pdf](http://documents.epo.org/projects/babylon/eponet.nsf/0/2464E1CD907399E0C12572D50031B5DD/$File/global_patent_data_coverage.pdf)

6) テーブルと呼ばれる複数の表の項目を互いに関連付けて情報を管理するデータベース。

第1表 ファミリーで比較した出願人名の表記ゆれ

日本国特許庁の出願情報			欧州特許庁の出願情報			アメリカ合衆国特許商標庁の出願情報		
出願番号	出願人名	国コード	出願番号	出願人名	国コード	出願番号	出願人名	国コード
JP1998107888	KYOEI SETSUBI:KK	なし	*** 対応ファミリー出願なし ***			US1999292921	Kyoei Plumbing Co., Ltd.	JP
JP1990108079	SONII KK	なし	EP1991303693	Sony Corporation	JP	US1991690559	Sony Corporation	JP
JP1992108075	SHERU INTERN RISAACHI MAACHATSUPII BV	なし	EP1992201176	KRATON Polymers Research B.V.	NL	US1993084683	Shell Oil Company	US
JP1994104614	TOKYO EREKUTORON KK	なし	EP1995104818	TOKYO ELECTRON LIMITED	JP	US1996634675	Tokyo Electron Yamanashi Limited	JP
JP1994104615	TOKYO ELECTRON LTD	なし	EP1995104818	TOKYO ELECTRON LIMITED	JP	US1996634675	Tokyo Electron Yamanashi Limited	JP
JP1995110193	BURIJISUTON SUHOOTSU KK	なし	*** 対応ファミリー出願なし ***			US1995424196	Bridgestone Sports Co., Ltd.	JP
JP1992121932	HAIDERUBAAGU HARISU INC	なし	EP1995119776	HEIDELBERG HARRIS INC.	US	*** 対応ファミリー出願なし ***		
JP1990121968	SUMITOMO WIRING SYST LTD	なし	EP1991107512	SUMITOMO ELECTRIC INDUSTRIES, LIMITED	JP	US1991696503	Sumitomo Electric Industries, Ltd.	JP
JP1991121967	VICTOR CO OF JAPAN LTD	なし	*** 対応ファミリー出願なし ***			US1993068570	VICTOR COMPANY OF JAPAN, LTD.	JP
JP1998195296	INTERNATL BUSINESS MACH CORP <IBM>	なし	*** 対応ファミリー出願なし ***			US2002036462	International Business Machines Corporation	US

出所 PATSTAT 2007

このような表記ゆれは、出願人が出願毎に自身の名前の表記を統一していても生じ、またこのような現象はPATSTATのデータに限られるものではありません。

例えば、第1図はアメリカ合衆国特許商標庁の発行した特許公報ですが、発明者の欄を見ると、「Mehmet Z. Sengun」とあります。この出願のファミリーが日本に出願された場合、日本国特許庁は、カタカナで「メメント ゼット センガン」と表音表記された願書を受理し、第2図の通り公開特許公報として発行され、更に公開特許英文抄録誌として翻訳されますと、「MEMETTO Z

ゼット センガン」はアルファベットで表記されますから、第3図に記載のように「MEMETTO Z SENGAN」となります。そしてこの情報がPATSTATに反映されますので、同じファミリーであってもPATSTATのデータのうち、アメリカ合衆国特許商標庁によるものは、発明者が「Mehmet Z. Sengun」と記載され、他方、日本国特許庁によるものは、第3図の記載と同じく「MEMETTO Z SENGAN」となり、異なるつづりで表記されてしまうことになります。

英語を日本語に表音表記し、さらにもう一度、日本語を英語に表音表記するためにこのような齟齬をきたすようです。

なお、日本語をアルファベットで表記した場合、ロー

マ字のヘボン式、訓令式のいずれを採用するかでつづりに違いが生じる場合があります。例えば、日本語の「株式会社」の表記は、PATSTATのデータベース上は、KABUSHIKIKAISHAとKABUSIKIKAISHAと2通りのつづりが存在します。

(2) 翻訳

ファミリーであっても、各国特許庁の出願情報によるデータソースの違いから、出願人名の表記が微妙に統一されていない例があります。

特願平07-010020号（特開平08-205421号公報）の出願人は三菱電機株式会社であり、このファミリーを日米欧で確認すると、公開特許英文抄録誌（PAJ; Patent Abstracts of Japan）では出願人としてMITSUBISHI ELECTRIC CORPと記載され、PATSTATのデータベース上は、日本がMITSUBISHI ELECTRICとあります。これに対し、欧州特許第0724321号明細書にはMITSUBISHI DENKI KABUSHIKI KAISHAと記載され、PATSTAT上もMITSUBISHI DENKI、また、米国特許第5686819号明細書もMitsubishi Denki Kabushiki Kaishaと記載され、PATSTAT上もMITSUBISHI DENKIと入力されています。つまり、第4図のように、同じ出願人でありながら英語表記が

US006328407B1

(12) United States Patent
Sengun

(10) Patent No.: US 6,328,407 B1
(45) Date of Patent: Dec. 11, 2001

(54) METHOD AND APPARATUS OF PREWARMING A PRINTHEAD USING PREPULSES

(75) Inventor: Mehmet Z. Sengun, Rochester, NY (US)

(73) Assignee: Xerox Corporation, Stamford, CT (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: 09/234,114
(22) Filed: Jan. 19, 1999
(51) Int. Cl.: B41J 2/05
(52) U.S. Cl.: 347/17; 347/60
(58) Field of Search: 347/11, 14, 17, 347/60, 56, 57, 185

(56) References Cited
U.S. PATENT DOCUMENTS
4,463,359 7/1984 Ayata et al. 347/56

5,745,132 * 4/1998 Hinbayashi et al. 347/14

FOREIGN PATENT DOCUMENTS

0 300 634 A	1/1989	(EP)
0 434 367 A	6/1991	(EP)
0 496 525 A1	7/1992	(EP)
0 505 154 A2	9/1992	(EP)
63 004970 A	1/1988	(JP)
WO 90/10540	9/1990	(WO)
WO 90/10541	9/1990	(WO)

(57) ABSTRACT
A method and apparatus of warming a printhead with a heat sink using prepulsing is provided. A temperature of the printhead may be measured and then using predetermined data, an amount of prewarming of the printhead may be determined based on the measured temperature. The printhead may be prewarmed by prepulses based on the determined amount of prewarming prior to printing a swath of data. The swath of data may then be printed and the temperature of the printhead may again be measured to determine a subsequent temperature and amount of pre-

* cited by examiner
Primary Examiner—Craig A. Hallacher
(74) *Attorney, Agent, or Firm*—Oliff & Berridge, PLC

第1図 アメリカ合衆国特許商標庁の発行する特許公報の例
出所 アメリカ合衆国特許商標庁

(19) 日本国特許庁 (J P) **(12) 公開特許公報 (A)** **(11) 特許出願公開番号**
特開2000-203025
(P2000-203025A)

(43) 公開日 平成12年7月25日 (2000.7.25)

(51) Int.Cl.⁷	識別記号	F I	テマコード⁷(参考)
B 4 1 J	2/05	B 4 1 J 3/04	1 0 3 B
	2/01		1 0 1 Z
	2/205		1 0 3 X

審査請求 未請求 請求項の数 1 O L (全 12 頁)

(21) 出願番号 特願2000-9110(P2000-9110)	(71) 出願人 590000798 ゼロックス コーポレイション XEROX CORPORATION アメリカ合衆国 06904-1600 コネティ カット州・スタンフォード・ロング リッ チ ロード・800
(22) 出願日 平成12年1月18日 (2000.1.18)	(72) 発明者 メメット ゼット センガン アメリカ合衆国 ニューヨーク州 ロチェ スター サウス グッドマン ストリート 270 アパートメント 301
(31) 優先権主張番号 0 9 / 2 3 4 1 1 4	(74) 代理人 100075258 弁理士 吉田 研二 (外2名)
(32) 優先日 平成11年1月19日 (1999.1.19)	
(33) 優先権主張国 米国 (US)	

第2図 日本国特許庁の発行する公開特許公報の例
出所 独立行政法人工業所有権情報・研修館

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 2000-203025
(43) Date of publication of application : 25.07.2000

(51) Int.Cl.	B41J 2/05 B41J 2/01 B41J 2/205
(21) Application number : 2000-009110	(71) Applicant : XEROX CORP
(22) Date of filing : 18.01.2000	(72) Inventor : MEMETTO Z SENGAN
(30) Priority Priority number : 99 234114 Priority date : 19.01.1999 Priority country : US	

(54) METHOD AND APPARATUS FOR PREHEATING PRINT HEAD USING PREPULSE

(57) Abstract:
PROBLEM TO BE SOLVED: To optimize the quantity of liquid being discharged from a print head.
SOLUTION: Temperature of a print head is measured and preheating quantity thereof is determined based on the measured temperature. The print head is preheated with a prepulse based on the preheating quantity thus determined before a data is printed. Subsequently, a data is printed, temperature of the

第3図 公開特許英文抄録誌の例
出所 独立行政法人工業所有権情報・研修館

統一されていないこととなります。DENKIとELECTRICが同義であることを知らない外国のユーザは両者が実体として同じ出願人であると認識せずに統計処理してしまう恐れがあります。

{	JP8205421	MITSUBISHI ELECTRIC
	EP0724321	MITSUBISHI DENKI
	US5686819	MITSUBISHI DENKI

第4図

(3) 名義変更

日本国特許庁の場合、特許電子図書館において提供されている「特許・実用新案文献番号索引照会」や「公報テキスト検索」で照会できる情報は公報が発行された時点のデータです。したがって、出願人名も公報発行時点のものとなります。名義変更などの内容は、「経過情報(番号照会)」で調べることができます。

PATSTATでは、権利の移転による出願人(譲受人)の変更がなされた場合、データ上、どのような扱いがなされているのでしょうか。

{	JP2000203025	XEROX
	EP1022139	SAMSUNG ELECTRONIC
	US6328407	XEROX

第5図

特願2000-009110号(特開2000-203025号公報)の出願人名は出願当初はXeroxでしたが、特許電子図書館の「経過情報(番号照会)」によれば、2005年4月に出願人名義変更届が出され、承継人として三星電子株式会社、つまりSamsungに権利譲渡されています。このファミリーを日米欧で比較すると、第5図のように、PATSTATのデータベース上は、米国特許第6328407号明細書に相当するデータがXEROXのままのに対し、欧州特許第1022139号明細書に相当するデータではSAMSUNG ELECTRONICと書き換えられています。同じファミリーであっても、PATSTATの記録上は、三極特許庁間で出願人名が異なることがわかります⁷⁾。この出願のファミリーをesp@snetで確認したものが第6図です。出願人、発明者の表記はこのリストの中でも完全には統一されていません。

Family list		Back to JP2000203025
7 family members for: JP2000203025 Derived from 5 applications		
1	No title available	in my patents list <input type="checkbox"/>
Inventor:		Applicant:
EC:		IPC:
Publication info: DE60030604D D1 - 2006-10-26		
2	Ink jet printers	in my patents list <input type="checkbox"/>
Inventor: SENGUN MEHMET Z (US)		Applicant: SAMSUNG ELECTRONICS CO LTD (KR)
EC: B41J2/05D		IPC: B41J2/01; B41J2/05; B41J2/205 (+3)
Publication info: DE60030604T T2 - 2007-11-08		
3	Ink jet printers	in my patents list <input type="checkbox"/>
Inventor: SENGUN MEHMET Z (US)		Applicant: XEROX CORP (US)
EC: B41J2/05D		IPC: B41J2/01; B41J2/05; B41J2/205 (+4)
Publication info: EP1022139 A2 - 2000-07-26 EP1022139 A3 - 2001-01-31 EP1022139 B1 - 2006-09-13		
4	METHOD AND APPARATUS FOR PREHEATING PRINT HEAD USING PREPULSE	in my patents list <input type="checkbox"/>
Inventor: MEMETTO Z SENGAN		Applicant: XEROX CORP
EC: B41J2/05D		IPC: B41J2/01; B41J2/05; B41J2/205 (+6)
Publication info: JP2000203025 A - 2000-07-25		
5	Method and apparatus of prewarming a printhead using prepulses	in my patents list <input type="checkbox"/>
Inventor: SENGUN MEHMET Z (US)		Applicant: XEROX CORP (US)
EC: B41J2/05D		IPC: B41J2/01; B41J2/05; B41J2/205 (+4)
Publication info: US6328407 B1 - 2001-12-11		

第6図 esp@snetによるファミリーリスト
出所 欧州特許庁

7) 日本では2002年の法務省令改正以前は登記上、商号にアルファベットを使用することはできなかったが、省令改正後はアルファベットとアラビア数字、「&」、「|」、「-」、「.」といった記号を使用できるようになり、これを機に商号を変更した例が散見される。

5. PATSTATの名寄せ

表記は異なっても実体としては同一の出願人であり、別々にカウントすると特許統計上、無視できない不都合が生じる場合、そのような出願人名を住所情報や市販の企業情報を用いて特定し、同一出願人として一元管理する、いわゆる「名寄せ」を行います。

この際、いつの時点のデータとして名寄せを行うかは重要です。4. (3) で紹介したように、名義変更がなされている場合や企業の合併や株式交換、株式移転による統合、株式の取得や営業の譲受による買収、分割などにより、一つの出願であっても、日付けにより、出願人名が異なってしまう場合があるからです。

ところで、日本国特許庁では、第2図に記載されるように、公報上での出願人名の原語表記に加え、第2表に例示されるように、出願事務において、世界の特許庁に先駆けて出願人コードを付与しており、庁内事務において実務上の出願人名の名寄せの問題はすでに改善されています。しかし、各国特許庁でそうした取り組みがなされているケースは少なく、したがって、特許統計データベースを用いてデータを抽出する際には、予め住所情報、業種といった属性を用いて名寄せを行う必要があります⁸⁾。なお、名寄せは、表記が同一もしくは類似した名前であって、異なる属性、主体の企業名に誤って行うことのないよう、あくまでも実

体として同じ出願人に対してのみ行うため、まずは名寄せの対象となりそうな候補を特定することから始めます。

6. 名寄せの方法

PATSTATに含まれる出願人名のうち、VOLKSWAGENとして名寄せの対象となりそうな出願人名を抽出し、名寄せを行った例が第3表です。このように、名寄せの対象となりそうな候補を抽出するためには、以下のようなデータ・クレンジングを行います。

(1) アルファベット文字の整理

PATSTATに記憶されている出願人名の情報はアルファベットが用いられています。アルファベットは代表的なラテン文字(ローマ字)に加え、ギリシア文字やキリル文字なども含みます。

PATSTATには日本語や中国語、韓国語などのアジアの言語は含まれていませんが、言語固有の文字コードとして様々なアルファベットが用いられており、同じ出願人が英語で出願した場合と他のアルファベット言語で出願した場合では、両者の出願人名は表記上、互いに異なるものとなり、その場合はやはり名寄せのためにデータ・クレンジングが必要となります。

第2表 日本国特許庁の付与する出願人コードの例

出願人名	出願人コード	2000年の特許出願件数
日本電子株式会社	4271	205
日本電子精機株式会社	593008427	1
日本電工株式会社	391021765	21
日本電業工作株式会社	232287	21
日本電気インフォメーションテクノロジー株式会社	390000756	15
日本電気オフィスシステム株式会社	232058	3
日本電気データ機器株式会社	232025	8
日本電気テレコムシステム株式会社	232106	3
日本電気ファクトリエンジニアリング株式会社	232139	1
日本電気株式会社	4237	4548

出所 OECD調べ

8) フランス特許庁 (INPI) では、国内で登記される全ての企業が有するSIRENコード (Système Informatisé du Répertoire des Entreprises) を用いて、出願人名の名寄せを行い、特許統計データの経済学的分析に役立てている。

<http://www.inpi.fr/fr/l-inpi/observatoire-de-la-propriete-intellectuelle.html>

(4)文字コードの統一

コンピュータ上では、出願人名を含め、全ての文字情報は文字コードの並びとして扱われます。したがってアルファベットの文字コードを統一し、HTMLやSGMLのコードを取り除いて、ASCIIに置き換えることとなります。

第3表 名寄せの例(1)

名寄せ前の出願人名	名寄せ後の出願人名
AKTIENGESELLSCHAFT VOLKSWAGEN	VOLKSWAGEN AG
FUORUKUSUAAGENUERUKU AG	
FUORUKUSUWAAGEN AG	
V O L K S W A G E N AKTIENGESE	
V W AG	
VOLKSWAGEN	
VOLKSWAGEN A G	
VOLKSWAGEN AG	
VOLKSWAGEN AG VW	
VOLKSWAGEN AKTIENGESELLSCHAFT	
VW	
VW AG	
VW WOLFSBURG	
WOLFSBURG VW	
BRASI S A VOLKSWAGEN DO	VOLKSWAGEN BRASIL
BRASIL S A VOLKSWAGEN DO	
BRASIL SA VOLKSWAGEN	

出所 OECD特許統計タスクフォース調べ

7. データ・クレンジングの留意点

「AB ELECTRONIK GMBH」と「AB Elektronik GmbH」のように、同じつづりであっても、大文字のみで表記される場合と小文字が混在している場合があります。この場合、つづりが同じであるからといって、全ての文字を単純に大文字もしくは小文字に統一して、データ・クレンジングを行うと、まれに問題が生じる場合があります。同じつづりであっても大文字で始まる単語と小文字で始まる単語はCapitonymと呼ばれ、異なる意味を有する場合があります。このような例は、英語のPolish（ポーランドの）とpolish（磨く）、ドイツ語のMorgen（朝）とmorgen（明日）、フランス語のMars（火星）とmars（3月）が挙げられます。出願人名については、ドメイン名のように大文字、小

文字の区別を無視して良い場合がほとんどですが、大文字もしくは小文字に表記を全て統一するようなデータ・クレンジングを行うと失われてしまう情報もあります。

第4表 名寄せの例(2)

名寄せ前のつづり	名寄せ後のつづり
SYSTEMEN	SYSTEM
SYSTEMES	
SYSTEME	
SYSTEMS	
TECHNOLOGIES	TECHNOLOGY
TECHNOLOGIEN	
TECHNOLOGIE	
INDUSTRIELLES	INDUSTRY
INDUSTRIELLE	
INDUSTRIELE	
INDUSTRIELS	

出所 OECD特許統計タスクフォース調べ

8. 難しい名寄せ

データ・クレンジングをした結果、名寄せをすべき候補を抽出したら、それらの関連性を住所情報や市販の企業情報を用いて調べることになりますが、特許統計データベースは膨大なデータを含んでいることから、同一出願人として名寄せを行うか否かを、単純な検索機能を使って割り出すことは、困難です。

プログラムを利用したデータ処理が現実的な対応となりますが、単純に名寄せの判断ができないケースは多々あります。例えば、「AAE HOLDING」と「AAE TECHNOLOGY INTERNATIONAL」や、「VVBH DEUTSCHLAND GMBH」と「IBM DEUTSCHLAND GMBH」は、それぞれ企業情報に照らして、同じ出願人とみなして良いのか否かを何らかの重みをつけずに単純なプログラムによる文字列の比較だけで判断することは難しいと言えます。日本の出願人名の例ですと「Japan as represented by the president of the university of Tokyo」と「President of Tokyo University」はやはり日本になじみのない人には判断のつきかねるケースのようです。名寄せの判断に際しては住所情報や業種を参酌する、ファミリーを有する特許出願人であって、日

本国特許庁にも出願されている場合、日本国特許庁の出願人コードを用いて名寄せをする、などの工夫をしていますが、簡単に判別のつかないケースは存在します。

名寄せを行って良いと判断した場合は、出願人名を一元化するために、唯一のコードをふり、リレーショナルデータベースとして、データの結合や抽出を行なうことができるようにします。

9. PATSTATのあるべき姿

OECDでは、イノベーションをどう測定するのか、イノベーションについての基礎的なデータをどう蒐集するかが目下の課題となっています。

PATSTATについては、体系的な特許統計データベースとして、出願人情報と企業・事業所情報との接合を図り、特許と経済の関係を他の統計情報などと合わせて分析できるようにすれば、利用価値は更に広がるものと考えられます。

特許統計は研究開発成果やイノベーション活動の指標として知られており、特許統計データベースを整備することにより、イノベーション政策に向けて様々な情報をもたらし、底の深い議論のきっかけを提供してくれるものと期待されています。

反面、特許統計は尽きることのない問題を抱えています。

OECDの特許統計タスクフォースの取り組みに対して、日本国特許庁は企画調査課を中心に、普及支援課、国際課の協力のもと、強いリーダーシップを発揮してきました。これは日本国特許庁の実務上の経験や知見によるところが大きいものと考えます。

本稿において紹介した、特許統計データベースの出願人名の名寄せに関し、日本国特許庁の庁内事務手続きでは、世界の特許庁に先駆けて出願人コードを付与してきた歴史があります。担当官の話では、実務上の必要性からそのような運用を開始した、とのことでしたが、こうした事実を知るたびに、日本国特許庁の諸先輩方々の先見性と真摯さに驚きます。

PATSTATがユーザにとってより使いやすいものとなるよう、批判や助言に支えられながらそのあるべき

姿について、特許統計タスクフォースの場を利用して、議論して参りたいと思います。

最後に、本稿は、OECDや日本国特許庁としての見解を示したものではなく、記事に含まれるであろう誤りは私個人の責任であり、何らかの有益な情報が含まれているとすれば、それはご指導いただいた多くの方々によるものであることを申し添えて結びといたします。

(ご意見・ご質問は下記までお願いいたします)

岡崎 輝雄

電子メール：teruo.okazaki@gmail.com

Profile

岡崎 輝雄 (おかざき てるお)

【経歴】

平成9年4月 特許庁入庁 (審査第二部建築)
 平成13年4月 審査官 (特許審査第一部ナノ物理)
 平成15年8月 独立行政法人国際協力機構 (JICA) 経済開発部出向
 平成17年8月 審査官 (特許審査第一部ナノ物理)
 平成18年6月 経済協力開発機構 (OECD) 科学技術産業局出向

【著書】

CAPTURING NANOTECHNOLOGY'S CURRENT STATE OF DEVELOPMENT VIA ANALYSIS OF PATENTS
 OECD STI WORKING PAPER 2007/4
 Statistical Analysis of Science, Technology and Industry
 Masatsura Igami and Teruo Okazaki
 JT03227650
<http://www.OECD.org/data/OECD/6/9/38780655.pdf>

OECD Science, Technology and Industry Scoreboard 2007
 Economic Analysis and Statistics Division (EAS) of the OECD
 Directorate for Science, Technology and Industry (DSTI)
 共著
 ISBN 978-92-64-03788-5
<http://caliban.sourceOECD.org/vl=12746996/cl=19/nw=1/rpsv/sti2007/foreword.htm>
http://www.sherpatv.it/media/13880/C_2_media_13880_oggetto.pdf

Compendium of Patent Statistics 2007
 Economic Analysis and Statistics Division (EAS) of the OECD
 Directorate for Science, Technology and Industry (DSTI)
 共著
<http://www.OECD.org/data/OECD/5/19/37569377.pdf>