

NISTEP 企業名辞書 (Ver. 2019_1)
利用マニュアル

文部科学省科学技術・学術政策研究所
(2019年1月)

目次

はじめに	1
1. 公開するファイルの概要	3
1. 1 NISTEP 企業名辞書	3
(1) 企業名辞書に掲載する条件	4
(2) 企業の名称変遷の取り扱い	5
(3) 企業に関する情報	6
1. 2 外部データ接続テーブル	6
2. 公開するファイルの内容	7
2. 1 企業名辞書	7
(1) テーブルの構成	7
①リレーショナルデータベース型企業辞書	7
②単一テーブル型企業名辞書	8
(2) テーブルのフィールド定義	8
2. 2 外部データ接続テーブル	17
(1) IIP パテントデータベースとの接続用	17
(2) NISTEP 大学・公的機関名辞書との接続用	17
(3) 米国特許との接続用	17
(4) 日本の会社データ 4 万社データとの接続用	18
3. 外部データベース	19
3. 1 IIP パテントデータベース	19
3. 2 NISTEP 大学・公的機関名辞書	19
3. 3 米国特許の出願人 (Applicant 又は Assignee) に記された国内営利企業名英語表記揺れテーブル	19
3. 4 日本の会社データ 4 万社	19
4. 公開ファイルのダウンロードとデータベースの構築	20
4. 1 ダウンロード	20
4. 2 データベースの構築	20
5. 企業名辞書の作成と名寄せ	21
5. 1 企業名辞書の作成	21
5. 2 企業名の変遷を考慮した名寄せ	22
【Appendix】	24
(1) 沿革 (history) を表現するコード	24
(2) 米国特許との接続	24
(2) - 1 接続対象米国特許	24
(2) - 2 日米特許の紐づけ	24
a) 国内特許から米国へパリ優先権を主張して米国へ出願した出願人情報の利用	24

b) 国際出願 (PCT 出願) から米国移行した特許情報の利用	24
c) 日米特許の紐づけ	25
(2) - 3 企業名辞書掲載企業と米国特許出願人 (applicant 又は assignee) の紐づけ	25
(2) - 4 接続テーブルの作成	25
(3) 企業名辞書を利用するための参考資料	25

【改訂履歴】

- (1) Ver. 2013.1 (2013年11月) : NISTEP 企業名辞書、各種接続テーブルとともに本書を初公開した。
- (2) Ver. 2014.1 (2014年6月) : 以下の改訂を行った。
- ・ NISTEP 企業名辞書;重複企業削除、誤記修正等のマイナー改訂を行った。
 - ・ IIP パテントデータベースとの接続テーブル; テーブル構造上、iipdb20140417 との接続が不可となった旧接続テーブルを改訂し接続可能とした。
 - ・ 利用マニュアル;上記に関連する記述の改訂を行った。
- (3) Ver. 2014.2 (2014年11月) : 以下の追加・改訂を行った。
- ・ NISTEP 企業名辞書;
 - ・ 2013年1月から2014年4月の期間に上場した企業で Ver. 2014.1 に未掲載の企業を追加した。
 - ・ 変遷情報の掲載未了であった上場企業約 2,400 社に関し、情報の追加を行った。
 - ・ 以下の項目を追加し、各企業に関する情報を追加した。沿革年(設立・名称変更・合併)、企業英語名称(上場企業のみ)、所在地大字町丁目コード、過去の上場実績
 - ・ 掲載全企業に関して、名称変更、合併等 2014年4月現在の状況を反映し見直した。
 - ・ IIP パテントデータベースとの接続テーブル; NISTEP 企業名辞書 (Ver. 2014.2) と iipdb20140417 とを接続するテーブル可能とした。
 - ・ 利用マニュアル;上記に関連する記述の改訂を行った。
- (4) Ver. 2015.1 (2015年10月) : 以下の追加・改訂を行った。
- ・ NISTEP 企業名辞書;
 - ・ 従来と同様の単一テーブル型企業名辞書に加えて、リレーショナルデータベース型企業名辞書も作成した。
 - ・ IIP パテントデータベース (iipdb20140417) から、特許出願件数(累積 100 件超)の企業を再評価し、新たに 100 件超となった企業など企業名辞書に未掲載の企業を追加した。
 - ・ iipdb20140417 から、特許出願件数伸びを 3、5、7 年の各期間で評価し、企業名辞書に未掲載の伸びの大きい企業を追加した。
 - ・ 2014年4月から2015年3月の期間に上場した企業を追加した。
 - ・ 掲載全企業に関して、名称変更、合併等 2015年4月現在の状況を反映し見直した。
 - ・ IIP パテントデータベースとの接続テーブル; NISTEP 企業名辞書 (Ver. 2015.1) と IIP パテントデータベース (2015年版) とを接続するテーブルに追加・改訂した。
 - ・ 利用マニュアル;上記に関連する記述の改訂を行った。
- (5) Ver. 2016.1 (2016年8月) : 以下の追加・改訂を行った。
- ・ NISTEP 企業名辞書;
 - ・ NISTEP 大学・機関名辞書 (ver. 2015.1) に掲載された科学論文著者の所属企業のうち、NISTEP 企業名辞書 (Ver. 2015.1) に未掲載の企業を追加登録(約 2,000 社)した。
 - ・ IIP パテントデータベース (2015年版) から、特許出願件数(累積 100 件超)の企業を再評価し、新たに 100 件超となった企業を追加登録した。
 - ・ IIP パテントデータベース (2015年版) から、特許出願件数伸びを 3、5、7 年の各期間で再評価し、企業名辞書に未掲載の伸びの大きい企業を追加登録した。
 - ・ 2015年4月から2015年6月の期間に上場した企業を追加登録した。
 - ・ 掲載全企業に関して、名称変更、合併等の沿革を 2015年6月現在の状況を反映し見直した。

- ・外部データベースとの接続テーブル；
 - ・IIP パテントデータベース（2016 年版）と NISTEP 企業名辞書（Ver. 2016. 1）との接続情報を追加・改訂した。
 - ・新たに、NISTEP 大学・公的機関名辞書との接続テーブルを作成し、企業レベルで論文（機関 ID で接続）と特許（企業 id で接続）の両者に接続できる構造とした。
- ・利用マニュアル；上記に関連する説明の改訂を行った。

(6) Ver. 2018. 1（2018 年 1 月）：以下の追加・改訂を行った。

- ・NISTEP 企業名辞書；
 - ・企業名辞書登録企業について、外部データとの接続性向上を目的として法人番号の情報を追加した。
 - ・以下の企業状況を 2016 年 6 月現在の状況を基本に確認し、変更のある場合情報の追加、改訂を行った。
 - ・企業名称が変更された場合
 - ・統合・再編等が行われた場合
 - ・株式上場（新規・再）、又は廃止した場合
 - ・倒産、清算等により会社解散が行われた場合
 - ・企業の連結関係に変化がある場合
 - ・その他、所在地、企業HP、規模、業種等の企業情報に変更がある場合
 - ・IIP パテントデータベース（2015 年版）から、沿革を考慮した企業名寄せを再評価し、未掲載であった特許出願件数（累積 100 件超）の企業を追加登録した。
 - ・IIP パテントデータベース（2015 年版）から、特許出願件数伸びを 3、5、7 年の各期間で再評価し、企業名辞書に未掲載の伸びの大きい企業を追加登録した。
 - ・2015 年 7 月から 2016 年 6 月の期間に上場した企業を追加登録した。
 - ・NISTEP 大学・機関名辞書（ver. 2016. 1）に掲載された科学論文著者の所属企業のうち、NISTEP 企業名辞書（Ver. 2016. 1）に未掲載の企業を追加登録した。
- ・外部データベースとの接続テーブル；
 - ・IIP パテントデータベース（2015 年版）と NISTEP 企業名辞書（Ver. 2018. 1）との接続情報を追加・改訂した。
 - ・NISTEP 大学・公的機関名辞書 ver. 2016. 1 との接続テーブルを更新した。
- ・利用マニュアル；上記に関連する説明の改訂を行った。

(7) Ver. 2018. 11（2018 年 4 月）：以下の改訂追加公開を行った。

- ・NISTEP 企業名辞書（Ver. 2018. 1）と IIP パテントデータベース（2015 年版）との接続テーブルについて、誤記、記載漏れ等のマイナー改訂を行った。
- ・米国特許との接続用テーブル及び米国特許の出願人に記された国内営利企業名英語表記ゆれテーブルの新規公開を行った。

(8) Ver. 2019_1（2019 年 1 月）：以下の追加・改訂を行った。

- ・NISTEP 企業名辞書；
 - ・以下の企業状況を 2017 年 10 月現在の状況を基本に確認し、変更のある場合情報の追加、改訂を行った。
 - ・企業名称が変更された場合

- ・ 統合・再編等が行われた場合
- ・ 株式上場（新規・再）、又は廃止した場合
- ・ 倒産、清算等により会社解散が行われた場合
- ・ 企業の連結関係に変化がある場合
- ・ その他、所在地、企業HP、規模、業種等の企業情報に変更がある場合
- ・ IIP パテントデータベース（2017年版）から、沿革を考慮した企業名寄せを再評価し、未掲載であった特許出願件数（累積100件超）の企業を追加登録した。
- ・ IIP パテントデータベース（2017年版）から、特許出願件数伸びを3、5、7年の各期間で再評価し、企業名辞書に未掲載の伸びの大きい企業を追加登録した。
- ・ 2016年7月から2017年10月の期間に上場した企業を追加登録した。
- ・ NISTEP 大学・機関名辞書（ver.2018.1）に掲載された科学論文著者の所属企業のうち、NISTEP 企業名辞書（Ver.2018.11）に未掲載の企業を追加登録した。
- ・ 意匠又は商標の登録件数の多い企業を追加登録した。
- ・ 外部データベースとの接続テーブル；
 - ・ IIP パテントデータベース（2017年版）との接続を前提に接続情報を追加・改訂した。
 - ・ NISTEP 大学・公的機関名辞書 ver.2018.1 との接続テーブルを更新した。
- ・ 利用マニュアル；上記に関連する説明の改訂を行った。

はじめに

文部科学省科学技術・学術政策研究所（以下、NISTEP と呼ぶ）は、NISTEP 企業名辞書、および、同辞書と外部データとを繋ぐ接続テーブルの公開を行う。

これらファイルの公開は、本文書と同じ web ページに掲載する「産業の研究開発に関するデータ整備について」に記したように、産業セクターの研究開発やイノベーションに関する分析・研究に用いるデータベースの構築に利用し、その分析・研究成果を通じて科学技術イノベーション政策の形成に貢献することを目的とする。

なお、以下の【公開するファイル】枠内に記したファイルの利用に際して、「NISTEP 企業名辞書」はクリエイティブ・コモンズ・ライセンス（CC ライセンス）の「表示-継承」を、「外部データ接続テーブル」は「表示-非営利」を適用する。

CC ライセンスの詳細は、<http://creativecommons.jp/licenses/> を参照願いたい。

また、表示するクレジットは下記の通りである。



（表示-継承）

又は、



（表示-非営利）

原作者名：文部科学省科学技術・学術政策研究所

作品タイトル：産業の研究開発に関するデータ

URL：<http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure>

【公開するファイル】

I. NISTEP 企業名辞書

- ・ RDB 型 NISTEP 企業名辞書 - Ver. 2019_1
タブ区切り txt 形式の構成テーブル一式【comp_name_dic_rdb_ver2019_1.zip】 **New**
- ・ 単一テーブル型 NISTEP 企業名辞書 - Ver. 2019_1
excel 形式【comp_name_dic_ver2019_1.xlsx】 **New**

II. 外部データ接続テーブル [NISTEP 企業名辞書と以下の外部データとの接続テーブル]

- ・ IIP パテントデータベース¹との接続用 - Ver. 2019_1
txt (タブ区切り) 形式【ct_comp_name_dic_vs_iip_ver2019_1.txt】 **New**
- ・ NISTEP 大学・公的機関名辞書との接続用 - Ver. 2019_1
txt (タブ区切り) 形式【ct_comp_name_dic_vs_organ_name_dic_ver2019_1.txt】 **New**
- ・ 日本の会社データ 4 万社²との接続用 - Ver. 2013.1
csv (カンマ区切り) 形式【ct_dic_nkdop_ver2013_1.csv】
- ・ 米国特許との接続用 - Ver. 2018.1
txt (タブ区切り) 形式【ct_comp_name_dic_vs_USPT_ver2018_1.zip】

< 続く >

¹ IIP パテントデータベースは、一般財団法人知的財産研究所より公開される特許庁の整理標準化データをもとに特許統計分析用に開発されたデータベースである。
IIP パテントデータベースの入手は同研究所のホームページより行い、定められた利用規約に従って利用願いたい。（<http://www.iip.or.jp/>）

² 日本の会社データ 4 万社は、市販データベースである。必要に応じて、発売社（東洋経済新報社）に確認願いたい。

< 続き >

Ⅲ. 表記揺れテーブル

- ・ 米国特許の出願人に記された国内営利企業名英語表記揺れテーブル
txt (タブ区切り) 形式【Notation Variations table of Corporate applicants_ver2018_1.zip】

【補足】

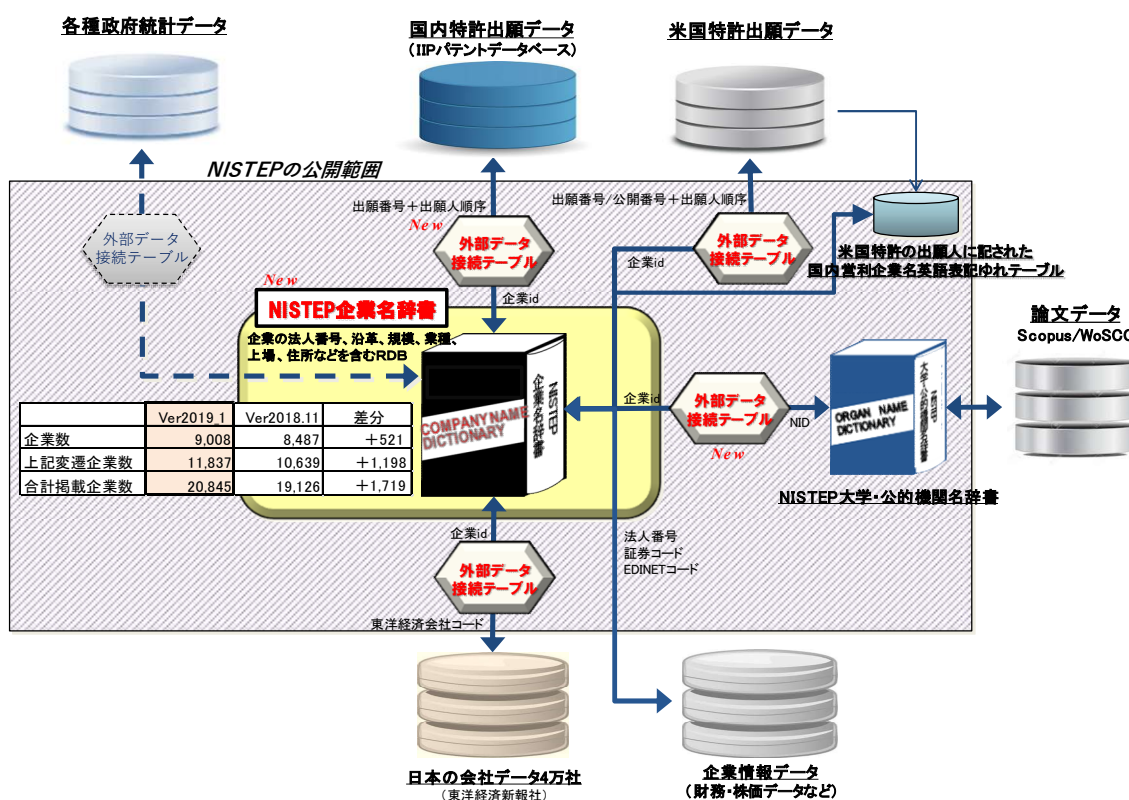
- ・ 【】内はファイル名
- ・ **New** の付けられたファイルは新規に作成 (又は改訂) され公開するファイルである。
- ・ txt ファイルのエンコード形式は UTF-8 BOM なし である。
- ・ ファイルの圧縮は zip 形式で行っている
- ・ Excel 形式の NISTEP 企業名辞書は、RDB 化された辞書情報を単一テーブルで見通しよく利用したい利用者の利便性を考え提供するものである

1. 公開するファイルの概要

公開するファイルは、産業セクターのイノベーション分析・研究に用いるデータベースの構築に必要な「NISTEP 企業名辞書」を核とし「外部データ接続テーブル」を加えた図表 1 に示す範囲である。

データベースの構築は、MySQL 等のデータベース管理システムを用い、txt 形式で提供する各ファイルおよび外部データをインポートしてテーブルを構成し、所定のフィールド間を関連付けて RDB 化する。

図表 1 産業セクターのイノベーション分析・研究に用いるデータベース連携の全体図



1. 1 NISTEP 企業名辞書

NISTEP 企業名辞書（以下、「企業名辞書」と呼ぶ）は、産業セクターのイノベーション分析・研究に用いるデータベースの中心に位置付けられ、国内営利企業（以下、「企業」と呼ぶ）に関し分析するための基本情報の提供や特許情報を始めとする各種の外部データベースを接続するための橋渡しなどハブとしての役割を担う。論文データとは NISTEP 大学・公的機関名辞書を経由して接続が可能であり、企業単位で特許情報と論文情報とを接続して分析することができる。また、企業名の読み、本社所在地、業種など、外部データに含まれる数多くの企業から分析対象である企業を特定しデータ抽出するための支援、および、合併や企業名称の変遷を考慮したデータの収集など、正しく企業情報に関する辞書として活用できる。

図表 1 で、企業名辞書は一冊の辞書のイラストで単純化し表しているが、Ver. 2015.1 以降、17 のテーブルで構成するリレーショナルデータベース（以下、「RDB」と略す）

形式に改め公開している。

なお、全ての利用者がRDBの操作に慣れている訳ではなく、機能が削られても良いから一つのテーブルで企業情報を見通せるように欲しいというニーズもあることからエクセル形式のファイルも公開している。

(1) 企業名辞書に掲載する条件

企業名辞書に掲載する企業は、原則、次の5つの条件の何れかを満足する企業の論理和で構成する。

- ① 特許出願数累積 100 件以上
- ② 株式上場企業
- ③ 特許出願数の伸び率大
- ④ NISTEP 大学・公的機関名辞書掲載企業
- ⑤ 意匠・商標登録数

① 特許出願数累積 100 件以上

IIP パテントデータベースの 2017 年版により、1970 年以降の企業の変遷（名称変更、合併）を考慮した特許出願の集約を行い、累積出願数が 100 件を超えた企業を掲載している。ちなみに、100 件を閾値としているのは、別途実施した企業による特許出願数の分析結果から、100 件以上の出願実績を持つ企業の出願数合計が全企業の出願数の 90%以上を占めていること、また、この閾値を下げると、指数関数的に該当企業が増える一方で占有率は僅かしか上昇しないことから決定している。

② 株式上場企業

株式上場企業（以下、「上場企業」と呼ぶ）は、2012 年 1 月時点の全上場企業を基準として、以降 2017 年 10 月までに新規（又は再）上場した企業を追加掲載している。

大手や中堅の製造業各社は条件①の基準を満たす確率が高い一方、例えば金融・保険業種などは有力企業であっても条件①からは抽出されず、イノベーション分析適用を考慮した業種ごとの抽出網羅性を高めるためにこの条件を設けている。なお、その後上場廃止した企業は上場廃止情報を所定のテーブル/フィールドに追加し、辞書から当該企業の抹消は行わず、他の掲載企業と同様に企業情報の更新を行っている。

③ 特許出願数の伸び率大

近年起業し活躍するベンチャー企業など、条件①、②では取りこぼす可能性がある企業の抽出を目的とする。ここでは、1970 年以降の企業の変遷を考慮した年ごとの出願数を把握し、それらデータを用いて 3 年、5 年、7 年の各期間で 1 年ごと移動させた線形フィットを行い、大きな回帰係数（出願数増分）を持つ企業を抽出する。線形フィットの傾きの閾値は、それぞれ 8、21/5、75/28 としている。この閾値は、当初、傾きの上位社から掲載した経緯があり、その後、最下位社の傾きを閾値として採用したことによる。

④ NISTEP 大学・公的機関名辞書掲載企業

科学論文データを中心に個別機関レベルでの体系的な分析を行うための基盤となる大学

・公的機関名辞書 (<http://www.nistep.go.jp/research/scisip/randd-on-university>) に含まれる企業である。それらの企業は論文著者所属企業から抽出されたものである。

⑤ 意匠・商標登録数累積 100 件以上

NISTEP 意匠・商標データベース³にて、意匠・商標登録数累積 100 件以上の企業である。

その他、上記条件に該当しない企業として、次の事由による掲載企業がある。

⑥ 持株会社制移行に伴い設立された事業会社

前記条件に合致し掲載された企業が純粋持株会社に移行し、同時に事業を承継するために新規に設立された会社である。ある事業会社が〇〇ホールディングス等に名称変更して純粋持株会社に移行し、同時に旧名称を引き継いだ事業子会社が新設分割される場合（抜け殻方式）が数多くある。世間的には、それら新旧の同一名称の会社は区別せず同じ企業として扱われている。このため、企業名辞書では新設事業子会社を追加掲載し、法人登記に合わせて純粋持株会社となった同一名称の旧会社と別法人として企業番号を付与している。

⑦ 一部事業の譲渡に伴い設立された会社

前記①～⑤の条件に合致し掲載された企業がその事業の一部を承継させるために新たに設立した会社。

⑧ 名称変更又は吸収合併した企業が登録事由に該当

名称変更は、前記④の条件（科学論文の著者所属企業）にのみ合致し掲載された企業が名称変更された場合における変更後の企業名称の掲載事由である。

吸収合併は、①～⑤の掲載条件に合致しない企業が、①～⑤の条件に合致した企業を吸収合併した場合の掲載事由である。

⑨ その他

企業名辞書作成初期における特許出願企業の名寄せ過程で紛れ込んだ同名異企業で、その後の精査で分離したものなどが少数存在する。

(2) 企業の名称変遷の取り扱い

企業は、名称の変更・合併等による企業形態の変化、清算・倒産などが日常的に起こっており、データ分析において、設立から現在までの一貫通貫した取り扱いに困難が伴うことが多い。また、同名異企業の存在により、異企業のデータを取り違えて使う恐れや、逆に、本来同一企業のデータと判別されるべきデータが異なる企業のものとして扱われてしまう恐れもある。

このような障害を排除するために、企業名辞書と連携する外部データは、現状の企業名称のみならず、沿革に基づいた旧名称や所在地等を考慮した名寄せを行い、可能な限り正

³ NISTEP で公開する NISTEP 意匠データベース及び NISTEP 商標データベースを利用した。このデータベースには 2000 年から 2013 年までに登録された意匠・商標に関するデータが収録されている。
<http://www.nistep.go.jp/archives/27214>

確に情報連携させるよう配慮している。

この前提として、企業名辞書には企業の沿革に沿った旧企業名称も掲載し、外部データの旧企業名称時代の情報は企業名辞書の旧企業名称に紐付けられるようにして時代ごとの取り扱いをできるようにしている。逆に、旧企業名称時代から現在までの情報を一括して取り扱いたい場合は、旧企業名称時代から現在名称まで共通の id (沿革 id) を用いて取り扱えるようにするなど、分析要求にフレキシブルに対応できるよう配慮した。

(3) 企業に関する情報

企業名辞書は、企業の主業に関する産業分類として、日本標準産業分類 (Japan Standard Industrial Classification) 及び証券コード協議会業種のほか、企業規模、沿革、連結関係、所在地、URL などの企業情報を含んだテーブルで構成している。

但し、財務データや株価データなどは、EDINET 等を通じた有価証券報告書からのデータ収集が比較的容易であることや、市販データも存在することから、それらデータの収集・公開は行っていない。市販データ等と接続する場合は、法人番号、証券コード及び EDINET コード情報を利用されたい。

1. 2 外部データ接続テーブル

外部データ接続テーブル (以下、「接続テーブル」と呼ぶ) は、企業名辞書に登録した企業と特許データ等の外部データに収められた当該企業 (特許の場合は出願企業) との接続を実現するテーブルであり、データ間の架け橋となる機能を有する。

また、接続テーブルには、外部データに含まれる企業名称の表記揺れを吸収する役割も持たせている。例えば、IIP パテントデータベースの出願人テーブルに記載された古い年代の出願人表記には、カタカナや OCR 読み取りに起因すると推測される多くの誤記や異称などが存在する。データ間の同一企業情報を接続する作業過程において、表記は揺れているが同一企業と判定できる企業は可能な限り接続テーブルに接続情報を含めている。データ間の同一企業の判別には、原則として企業名称、法人格コード (前株後株等をコード化)、住所コード (市区町村レベル) の三つの要素を使用するが、ここでは目視作業により機械的なマッチングでは不可能な接続を補完している。これにより、企業名辞書の正式企業名称から、外部データの表記揺れした企業のデータ取得を可能としている。

2. 公開するファイルの内容

ここでは、公開するファイルのテーブル構成、及び各テーブルのフィールドについて説明する。

2. 1 企業名辞書

(1) テーブルの構成

企業名辞書は、単一テーブル型の企業名辞書（comp_name_dic_ver2019_1.xlsx）と、データの共通性や管理に配慮し正規化し複数のテーブル構成としたリレーショナルデータベース型の企業辞書（comp_name_dic_rdb_ver2019_1.zip）の2種類を提供する。

① リレーショナルデータベース型企業辞書

企業名辞書は、2015年にリリースしたVer.2015.1以降は、図表2に示す17のテーブルで構成するリレーショナルデータベース構造で提供する。（Entity Relationship Diagramは図表25/P20参照）

これはデータの正規化によるメンテナンス性の向上や企業データをパネル化できるようにする要望が大きかったことによる。

各テーブルの概要は図表2に一覧としてまとめておく。なお、図表中の論理名は利用者に意味が通じる名称であるのに対して、物理名はコンピューターが識別する名称を意味している。

図表2 企業名辞書の構成テーブル

番号	テーブル名称		概要
	論理名	物理名	
1	企業名辞書メインテーブル	1_comp_name_main_TBL	企業名、企業id等のメインの情報、およびパネルデータとして整備をする必要がなく、最新の情報のみ保持すればよいデータを保管
2	沿革テーブル	2_comp_history_TBL	名称変更や吸収合併などの事象が発生した際に発生した年、事象の種類を保管
3	所在地テーブル	3_address_TBL	企業の所在地に関する情報を保管 本社、本店、移転など複数の住所情報の保管、パネル化が可能
4	企業規模テーブル	4_comp_size_TBL	資本金、従業員数、中小企業基本法による企業規模情報を保管 規模測定年ごとのパネル化が可能
5	業種（証券コード協会）テーブル	5_ind_class_tse_TBL	証券コード協会の定める当該企業の業種区分を保管 属する分類が変更された際のパネル化が可能
6	業種（日本標準産業分類）テーブル	6_ind_class_jsic_TBL	主業の日本標準産業分類を保管 属する分類が変更された際のパネル化が可能
7	EDINETコードテーブル	7_edinet_code_TBL	EDINETのコードを保管 コードが変更された際のパネル化が可能
8	証券コードテーブル	8_sec_code_TBL	証券コードを保管 コードが変更された際のパネル化が可能
9	連結企業テーブル	9_consolidate_TBL	連結子会社である場合の親企業情報を保管 連結関係の変化のパネル化が可能
10	データ登録条件マスターテーブル	10_reg_reason_TBL	企業が企業名辞書に登録された理由に関するマスターテーブル
11	企業名称使用開始事象マスターテーブル	21_use_name_start_event_TBL	新設、旧名称からの名称変更等、企業名称の使用が開始された場合の使用開始事象に関するマスターテーブル

番号	テーブル名称		概要
	論理名	物理名	
12	企業名称使用終了事象マスターテーブル	22_use_name_end_event_MTBL	名称変更、吸収合併など、企業名称の使用が終了した場合の使用終了事象に関するマスターテーブル
13	事業所区分マスターテーブル	31_office_class_MTBL	住所情報の本社、本店、事業所等を判定するためのマスターテーブル
14	業種（証券コード協会）マスターテーブル	51_tse_MTBL	証券コード協議会の定める業種区分に関するマスターテーブル
15	業種（日本標準産業分類）マスターテーブル	61_jsic_MTBL	日本標準産業分類に関するマスターテーブル 平成 25 年 10 月改定・平成 26 年 4 月 1 日施行に準拠
16	企業連結事象発生マスターテーブル	91_consolidate1_MTBL	連結事象が発生した場合の発生理由(子会社化等)に関するマスターテーブル
17	企業連結事象終了マスターテーブル	92_consolidate2_MTBL	連結事象が終了した場合の発生理由(他社の子会社となった、独立した等)

②単一テーブル型企业名辞書

企業名辞書をリレーショナルデータベース化したことによる副作用もあり、企業ごとのデータ参照に手間がかかり辞書として如何なものかという意見が存在することや、特に SQL のスキルを持たない利用者にとっては扱い辛いことも確かである。このため、一部の情報に制約ができるが、従来と同様にエクセル形式の単一テーブル型の企業名辞書を提供する。

(2)テーブルのフィールド定義

図表 2 の企業名辞書の構成テーブルに関して、フィールド（項目）を図表 3～図表 19 に定義する。

なお、表中の Y/N は以下を意味する。

重複 Y=重複を許す、N=重複を許さない

NULL Y=NULL を許す、N=NULL を許さない

主キー データを一意に決定する事が出来る項目に Y

外部キー 外部データベースとの接続に利用するキーの場合、当該テーブルの対応

図表 3 企業名辞書メインテーブルのフィールド定義

企業名辞書メインテーブル [1_comp_name_main_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	説明
論理名	物理名						
企業番号	comp_id	数値 (整数)	N	N	Y		企業（企業名称ごと）に固有に付与した番号
法人番号	corp_num	数値 (整数)	Y	Y			国税庁により付与された法人番号
沿革番号	history_id	数値 (整数)	Y	N			同一企業の変遷レコードをグループ化して扱うための番号
企業名称	comp_name	文字列	Y	Y			企業の名称（変遷名称も含む）
ふりがな	read	文字列	Y	Y			上記企業名称のふりがな
法人格コード	comp_code	文字列	Y	Y			企業の法人格を表すコード（下表参照）
英語名称	e_name	文字列	Y	Y			企業の英語名称

企業名辞書メインテーブル [1_comp_name_main_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	説明
論理名	物理名						
URL	url	文字列	Y	Y			企業のウェブページのURL
データ登録理由番号	reg_reason_id	数値(整数)	Y	Y		データ登録理由マスターテーブルの理由番号	当該企業の辞書掲載条件
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

法人格コード	例
KB1	株式会社〇〇〇
KB2	〇〇〇株式会社
KB3	〇〇〇株式会社□□□
YG1	有限会社〇〇〇
YG2	〇〇〇有限会社
GD1	合同会社〇〇〇
GD2	〇〇〇合同会社
SG1	相互会社〇〇〇
SG2	〇〇〇相互会社
GS1	合資会社〇〇〇
GS2	〇〇〇合資会社
GM1	合名会社〇〇〇
GM2	〇〇〇合名会社

図表 4 沿革テーブルのフィールド定義

沿革テーブル [2_comp_history_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	説明
論理名	物理名						
企業番号	comp_id	数値(整数)	N	N	Y	企業名辞書メインテーブルの企業番号	企業(企業名称ごと)に固有に付与した番号
企業名称使用開始年	uns_year	YEAR	Y	Y			企業名称の使用を開始した年
企業名称使用開始事象id	uns_event_id	数値(整数)	Y	Y		企業名称使用開始事象マスターテーブルの事象番号	企業名称の使用を開始した事由
事象発生前企業id	pre_comp_id	数値(整数)	N	Y		企業名辞書メインテーブルの企業番号	事象発生前に名乗っていた企業名称につけられた企業番号
企業名称使用終了年	une_year	YEAR	Y	Y			企業名称の使用が終了した年
企業名称使用終了事象id	une_event_id	数値(整数)	Y	Y		企業名称使用終了事象マスターテーブルの事象番号	企業名称の使用を終了した事由
事象発生後企業id	post_comp_id	数値(整数)	Y	Y		企業名辞書メインテーブルの企業番号	事象終了後の新しい名称につけられた企業番号
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 5 所在地テーブルのフィールド定義

所在地テーブル [3_address_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値 (整数)	Y	N	Y	企業名辞書メインテーブルの企業番号	企業 (企業名称ごと) に固有に付与した番号
所在地利用開始年	ads_year	YEAR	Y	Y			所在地で活動を開始した年
所在地利用終了年	ade_year	YEAR	Y	Y			所在地で活動を終了した年
事業所区分コード	office_code	数値 (整数)	Y	N	Y	事業所区分マスターテーブルの区分コード	本店、本社、事業所等の区分コード
所在地	address	文字列	Y	N	Y		所在地住所
都道府県コード	pref_code	数値 (2桁整数)	Y	Y			所在地の都道府県を示すコード
地方自治体コード	city_code	数値 (5桁整数)	Y	Y			所在地の市区町村レベルまでの住所コード
住所コード	add_code	数値 (12桁整数)	Y	Y			所在地の大字・町丁目を示す国道交通省が定める住所コード
緯度	latitude	数値 (実数)	Y	Y			所在地の北緯
経度	longitude	数値 (実数)	Y	Y			所在地の東経
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 6 企業規模テーブルのフィールド定義

企業規模テーブル [4_comp_size_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値 (整数)	Y	N	Y	企業名辞書メインテーブルの企業番号	企業 (企業名称ごと) に固有に付与した番号
企業規模測定年	judg_year	YEAR	Y	N	Y		企業規模を確認した年
中小企業基本法	comp_size_law	文字列	Y	Y			中小企業基本法に準拠し判定した企業規模
資本金階級	comp_size_cap	文字列	Y	Y			資本金の該当階級
従業員数階級	comp_size_emp	文字列	Y	Y			従業員数の該当階級
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

企業規模 (中小企業基本法) の区分

大企業者
 中小企業者
 小規模企業者

資本金階級
 100 万円未満
 100 万円以上

1000 万円以上
 2000 万円以上
 5000 万円以上
 1 億円以上
 10 億円以上

従業員階級

5 人未満
 5～29 人
 30～99 人
 100～299 人
 300～999 人
 1,000～4,999 人
 5,000 人以上

図表 7 業種（証券コード協会）テーブルのフィールド定義

業種（証券コード協会）テーブル [5_ind_class_tse_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値（整数）	Y	N	Y	企業名辞書メインテーブルの企業番号	企業（企業名称ごと）に固有に付与した番号
業種分類開始年	inds_year	YEAR	Y	Y			証券コード協会の業種分類の確認初年
業種分類終了年	inde_year	YEAR	Y	Y			証券コード協会の業種分類の確認最終年
業種分類コード	ind_code	数値（4桁整数）	Y	N	Y	業種（証券コード協会）マスターテーブルの分類コード	証券コード協会の分類該当業種
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 8 業種（日本標準産業分類）テーブルのフィールド定義

業種（日本標準産業分類）テーブル [6_ind_class_jsic_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値（整数）	Y	N	Y	企業名辞書メインテーブルの企業番号	企業に固有の id
JSIC 開始年	jsics_year	YEAR	Y	Y			JSIC 分類を確認した最初の年
JSIC 終了年	jsice_year	YEAR	Y	Y			JSIC 分類を確認した最後の年
JSIC 分類番号	jsic_code	文字列	Y	N	Y	業種（日本標準産業分類）マスターテーブルの JSIC 分類番号	JSIC 分類該当業種
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 9 EDINET コードテーブルのフィールド定義

EDINET コードテーブル [7_edinet_code_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値(整数)	Y	Y	Y	企業名辞書メインテーブルの企業番号	企業(企業名称ごと)に固有に付与した番号
EDINETコード確認年	check_year	YEAR	Y	Y	Y		EDINETコードを確認した年
EDINETコード	edinet_code	文字列	Y	Y			電子開示システムEDINETにおけるコード
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 10 証券コードテーブルのフィールド定義

証券コードテーブル [8_sec_code_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値(整数)	Y	Y	Y	企業名辞書メインテーブルの企業番号	企業(企業名称ごと)に固有に付与した番号
証券コード	sec_code	数値(4桁整数)	Y	Y	Y		証券コード協議会が付与する証券コード
上場市場	stock_market	文字列	Y	Y			企業が上場した東証等の市場
上場日	listed_date	年月	Y	Y			企業が上場した日
上場廃止日	delisted_date	年月	Y	Y			企業が上場を廃止した日
ISINコード	isin_code	文字列(12桁)	Y	Y			国際的に統一された証券コード
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 11 連結企業テーブルのフィールド定義

連結企業テーブル [9_consolidate_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
企業番号	comp_id	数値(整数)	Y	Y	Y	企業名辞書メインテーブルの企業 id	企業(企業名称ごと)に固有に付与した番号
連結事象発生年	cons_year	YEAR	Y	Y			連結事象が発生した年
連結事象発生事象 id	cons_id	数値(整数)	Y	Y		連結事象発生マスターテーブルの事象 id	連結事象発生前の当該企業の連結状況
連結事象発生前連結企業 id	post_parent_comp_id	数値(整数)	Y	Y		企業名辞書メインテーブルの企業 id	連結事象発生前の親会社の企業 id
連結先企業 id	parent_comp_id	数値(整数)	Y	N	Y	企業名辞書メインテーブルの企業 id	連結の親会社の企業 id
連結事象終了年	cone_year	YEAR	Y	Y			連結事象が終了した年
連結事象終了事象 id	cone_id	数値(整数)	Y	Y		連結事象終了マスターテーブルの事象 id	連結事象終了後の当該企業の連結状況

連結企業テーブル [9_consolidate_TBL]							
フィールド名		データ	重複	NULL	主キ	外部キー	詳細
連結事象 終了後連 結企業 id	new_paren t_compid	数値(整 数)	Y	Y		企業名辞書メインテーブ ルの企業 id	連結事象終了後の親会 社の企業 id
データ 登録日	reg_year	年月日	Y	N			データを本テーブルに 登録した日
データ 更新日	up_year	年月日	Y	N			既登録データの情報更 新した日

図表 12 データ登録条件マスターテーブルのフィールド定義

データ登録理由マスターテーブル [10_reg_reason_M_TBL]							
フィールド名		データ 型	重複	NULL	主キ ー	外部キー	詳細
論理名	物理名						
条件番号	re_id	数値(整 数)	N	N	Y	企業名辞書メインテーブ ルの企業 id	企業が企業名辞書に登 録された抽出条件を示 す id
登録条件	reason	文字列	N	Y			データが企業名辞書に 抽出登録された条件
データ 登録日	reg_date	年月日	Y	N			データを本テーブルに 登録した日
データ 更新日	up_date	年月日	Y	N			既登録データの情報更 新した日

条件番号	登録条件
1	特許出願件数
2	特許出願件数伸び
3	上場
4	大学公的機関名辞書
5	意匠または商標出願件数
6	持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
7	名称変更又は吸収合併した企業が登録事由に該当
12	特許出願件数・特許出願件数伸び
13	特許出願件数・上場
14	特許出願件数・大学公的機関名辞書
15	特許出願件数・意匠または商標出願件数
17	特許出願件数・特許出願件数伸び・名称変更又は吸収合併した企業が登録事由に該当
23	特許出願件数伸び・上場
24	特許出願件数伸び・大学公的機関名辞書
25	特許出願件数伸び・意匠または商標出願件数
26	特許出願件数伸び・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
27	特許出願件数伸び・名称変更又は吸収合併した企業が登録事由に該当
34	上場・大学公的機関名辞書
35	上場・意匠または商標出願件数
36	上場・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
45	大学公的機関名辞書・意匠または商標出願件数
47	大学公的機関名辞書・名称変更又は吸収合併した企業が登録事由に該当
56	意匠または商標出願件数・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
123	特許出願件数・特許出願件数伸び・上場
124	特許出願件数・特許出願件数伸び・大学公的機関名辞書
125	特許出願件数・特許出願件数伸び・意匠または商標出願件数
126	特許出願件数・特許出願件数伸び・持株会社制移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
134	特許出願件数・上場・大学公的機関名辞書

135	特許出願件数・上場・意匠または商標出願件数
145	特許出願件数・大学公的機関名辞書・意匠または商標出願件数
156	特許出願件数・意匠または商標出願件数・持株会社移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
234	特許出願件数伸び・上場・大学公的機関名辞書
235	特許出願件数伸び・上場・意匠または商標出願件数
245	特許出願件数伸び・大学公的機関名辞書・意匠または商標出願件数
345	上場・大学公的機関名辞書・意匠または商標出願件数
1234	特許出願件数・特許出願件数伸び・上場・大学公的機関名辞書
1235	特許出願件数・特許出願件数伸び・上場・意匠または商標出願件数
1245	特許出願件数・特許出願件数伸び・大学公的機関名辞書・意匠または商標出願件数
1246	特許出願件数・特許出願件数伸び・大学公的機関名辞書・持株会社移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
1256	特許出願件数・特許出願件数伸び・意匠または商標出願件数・持株会社移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
1345	特許出願件数・上場・大学公的機関名辞書・意匠または商標出願件数
2345	特許出願件数伸び・上場・大学公的機関名辞書・意匠または商標出願件数
12345	特許出願件数・特許出願件数伸び・上場・大学公的機関名辞書・意匠または商標出願件数
12356	特許出願件数・特許出願件数伸び・上場・意匠または商標出願件数・持株会社移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社
12456	特許出願件数・特許出願件数伸び・大学公的機関名辞書・意匠または商標出願件数・持株会社移行に伴い設立された事業会社又は一部事業の譲渡に伴い設立された会社

図表 13 企業名称使用開始事象マスターテーブルのフィールド定義

企業名称使用開始事象マスターテーブル [21_use_name_start_event_M_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
事象番号	uns_event_id	数値(整数)	N	N	Y		企業名称使用開始事象の id (下表参照)
事象概要	uns_event	文字列	N	Y			企業名称使用開始事象の概要 (下表参照)
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

事象 id	事象概要
1	設立
2	名称変更

図表 14 企業名称使用終了事象マスターテーブルのフィールド定義

企業名称使用終了事象マスターテーブル [22_use_name_end_event_M_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
事象番号	une_event_id	数値(整数)	N	N	Y		企業名称使用終了事象の id (下表参照)
事象概要	une_event	文字列	N	Y			企業名称使用終了事象の概要 (下表参照)
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

事象 id	事象概要
1	現存
2	名称変更

3	合併
4	破産
5	清算

図表 15 事業所区分マスターテーブルのフィールド定義

事業所区分マスターテーブル [31_office_class_M_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
事業所区分コード	office_code	数値（整数）	N	N	Y		本店、本社、事業所等の区分コード（下表参照）
事業所区分	class	文字列	N	N			本店、本社、その他の事業所等の区分（下表参照）
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

事業所区分コード	概要
1	本社
2	本店
3	研究所
4	事業所
5	その他

図表 16 業種（証券コード協会）マスターテーブルのフィールド定義

業種（証券コード協会）マスターテーブル [51_tse_M_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
業種コード	class_code	数値（4桁整数）	Y	N	Y		証券コード協会の業種分類によるコード
版	tse_ver	年月日	Y	N	Y		証券コード協会の業種分類の版
大分類	l_class	文字列	Y	N			証券コード協会の業種大分類
小分類	s_class	文字列	Y	N			証券コード協会の業種小分類
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 17 業種（日本標準産業分類）マスターテーブルのフィールド定義

業種（日本標準産業分類）マスターテーブル [61_jsic_M_TBL]							
フィールド名		データ型	重複	NULL	主キー	外部キー	詳細
論理名	物理名						
JSIC 分類番号	jsic_code	文字列	Y	N	Y		日本標準産業分類の分類番号-JISCO〇〇〇〇
JSIC 版	jsic_ver	日付	Y	N	Y		日本標準作業分類の版（改訂日）
JSIC 大分類	jsic_l	文字列	Y	N			日本標準産業分類大分類
JSIC 中分類	jsic_m	文字列	Y	N			日本標準産業分類中分類
JSIC 小分類	jsic_s	文字列	Y	N			日本標準産業分類小分類

業種（日本標準産業分類）マスターテーブル [61_jsic_MTBL]							
フィールド名		データ型	重複	NULL	主キ	外部キー	詳細
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

図表 18 企業連結事象発生マスターテーブルのフィールド定義

企業連結事象発生マスターテーブル [91_consolidate1_MTBL]							
フィールド名		データ型	重複	NULL	主キ	外部キー	詳細
論理名	物理名						
事象番号	cons_id	数値（整数）	N	N	Y		連結事象発生事象の番号
事象概要	cons_event	文字列	N	Y			連結事象発生事象の概要
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

事象 id	事象概要
1	連結子会社として設立
2	名称変更以前から連結子会社であった
3	独立企業から連結子会社になった
4	他社の連結子会社から当社の連結子会社になった

図表 19 企業連結事象終了マスターテーブルのフィールド定義

企業連結事象終了マスターテーブル [92_consolidate2_MTBL]							
フィールド名		データ型	重複	NULL	主キ	外部キー	説明
論理名	物理名						
事象番号	cone_id	数値（整数）	N	N	Y		連結事象終了事象の番号
事象概要	cone_event	文字列	N	Y			連結事象終了事象の概要
データ登録日	reg_date	年月日	Y	N			データを本テーブルに登録した日
データ更新日	up_date	年月日	Y	N			既登録データの情報更新した日

事象 id	事象概要
1	現在も連結子会社である
2	名称変更
3	連結子会社から独立企業になった
4	当社の連結子会社から他社の連結子会社になった
5	連結先に合併されて解散した
6	連結先以外に合併されて解散した
7	合併以外の理由により解散した

2. 2 外部データ接続テーブル

(1) IIP パテントデータベースとの接続用

企業名辞書と外部データである IIP パテントデータベースを連携させるための接続テーブルであり、企業名辞書メインテーブルの企業 id と IIP パテントデータベースの出願人テーブルの ida_seq フィールドを関係付け接続する。

図表 20 接続テーブル（IIP パテントデータベースとの接続用）のフィールド定義

フィールド名		データ型	説明
論理名	物理名		
企業番号	comp_id	数値（整数）	企業（企業名称ごと）に固有に付与した番号
IIP パテント出願番号 + 記載順序	ida_seq	文字列	上記企業番号の企業が出願人である特許
出願日	adate	日付	特許の出願日

(2) NISTEP 大学・公的機関名辞書との接続用

企業名辞書と NISTEP 大学・公的機関名辞書、又は企業名辞書と科学論文-NISTEP 大学・公的機関名辞書対応テーブルと連携させるための接続テーブルであり、企業名辞書メインテーブルの企業 id と NISTEP 大学・公的機関名辞書の機関番号を関係付け接続する。

図表 21 接続テーブル（NISTEP 大学・公的機関名辞書との接続用）のフィールド定義

フィールド名		データ型	説明
論理名	物理名		
企業番号	comp_id	数値（整数）	企業（企業名称ごと）に固有に付与した番号
機関 id	n_id	文字列	NISTEP 大学・公的機関名辞書に掲載する機関の識別のために付与した番号
下部組織フラグ	low_org	数値（整数）	NISTEP 大学・公的機関名辞書に掲載する企業には、部署、社内カンパニー、企業立病院など企業の下部組織が一部含まれている。下部組織フラグはそれらを識別するためのフラグである。

フラグ番号	下部組織フラグ説明
1	部署
2	社内カンパニー
3	企業立病院

(3) 米国特許との接続用

企業名辞書メインテーブルと外部データである米国特許を連携させるための接続テーブルであり、企業名辞書メインテーブルの企業番号と米国特許出願番号又は米国特許公開番号にて接続を行う。

接続テーブル作成の詳細は、「【Appndix】(2) 米国特許との接続」を参照されたい。

図表 22 接続テーブル（米国特許との接続用）のフィールド定義

フィールド名		データ型	説明
論理名	物理名		
企業番号	comp_id	数値（整数）	企業（企業名称ごと）に固有に付与した番号
IIP パテント出願番号 + 記載順序	iip_ida	文字列	上記企業番号の企業が出願人である国内出願特許
米国特許（公開番号）	uspt_pub	文字列	上記企業番号の企業が出願人に含まれる米国出願特許
米国特許（出願番号）	uspt_app	文字列	上記企業番号の企業が出願人に含まれる米国出願特許

(4) 日本の会社データ 4 万社データとの接続用

企業名辞書メインテーブルと外部データである東洋経済新報社の「日本の会社データ 4 万社」を連携させるための接続テーブルであり、企業名辞書メインテーブルの企業番号と東洋経済企業コードを対応させ関係付けしている。

なお、日本の会社データ 4 万社は、2011 年 4 月版を対象に接続テーブルを作成している。その後、新版が発行されているが、東洋経済企業コードに変更がない限り、2011 年版で対応付けされた企業の範囲で接続可能である。

また、この接続テーブルにおける東洋経済企業コードの使用と公開は株式会社東洋経済新報社より許諾を受けている。本稿執筆時点では、許諾権の制約から新版対応の接続テーブルを作成する計画はない。

図表 23 接続テーブル（日本の会社データ 4 万社データとの接続用）のフィールド定義

フィールド名		データ型	説明
論理名	物理名		
企業番号	comp_id	数値（整数）	企業（企業名称ごと）に固有に付与した番号
東洋経済企業コード	code_tk	文字列	東洋経済新報社が企業の識別のために付与した独自の番号

3. 外部データベース

産業セクターのイノベーション分析・研究に用いるデータベースのうち、外部データに関しては次の3. 1及び3. 4を除きNISTEPの公開対象外である。

以下に、参考用として、外部データベースの簡単な説明を加えておく。

3. 1 IIP パテントデータベース

一般財団法人知的財産研究所（略称 IIP）が提供する特許データベースであり、特許庁の特許整理標準化データをリレーショナル型のデータベースに変換し、イノベーション研究に必要と思われる項目を集め公開している。

データベースは、出願テーブル、出願人テーブル、発明者テーブル等に構造化されている。企業名辞書との接続は、このうち、出願人テーブルと行う。

IIPのURL：<http://www.iip.or.jp/>

3. 2 NISTEP 大学・公的機関名辞書

研究活動を行っている我が国の機関（約1万9千機関）を掲載した機関名辞書であり、論文の著者企業も含まれている。企業名辞書はそれらの企業と接続でき、その結果、特許と論文を企業レベルで接続することができる。

<http://www.nistep.go.jp/research/scisip/randd-on-university>

3. 3 米国特許の出願人（Applicant 又は Assignee）に記された国内営利企業名英語表記揺れテーブル

企業名辞書と米国特許との接続情報の生成の中で整理した米国特許の出願人（Applicant 又は Assignee）に記された国内営利企業名英語表記揺れテーブルである。

図表 24 米国特許の出願人に記された国内営利企業名英語表記揺れテーブル
のフィールド定義

フィールド名		データ型	説明
論理名	物理名		
米国特許出願企業名	uspt_comp_name	文字列	米国特許出願に記載された企業名表記
記載出願件数順位	rank	数値（整数）	上記企業表記の出願件数順位
企業名称	comp_name	文字列	企業名辞書記載の企業名表記
法人格コード	comp_code	文字列	企業の法人格を表すコード
企業番号	comp_id	数値（整数）	企業（企業名称ごと）に固有に付与した番号
法人番号	corp_num	数値（整数）	国税庁により付与された法人番号
沿革番号	history_id	数値（整数）	同一企業の変遷レコードをグループ化して扱うための番. history_id=1の企業が最新企業となる

3. 4 日本の会社データ 4 万社

東洋経済新報社が独自調査で収集した上場会社および非上場会社の合計4万社弱の企業情報を収録した商業データベースである。

1社1レコード形式で、csvファイルで提供されている。

4. 公開ファイルのダウンロードとデータベースの構築

4.1 ダウンロード

公開したファイルは、任意のディレクトリにダウンロードし、圧縮形式のファイルは解凍する。

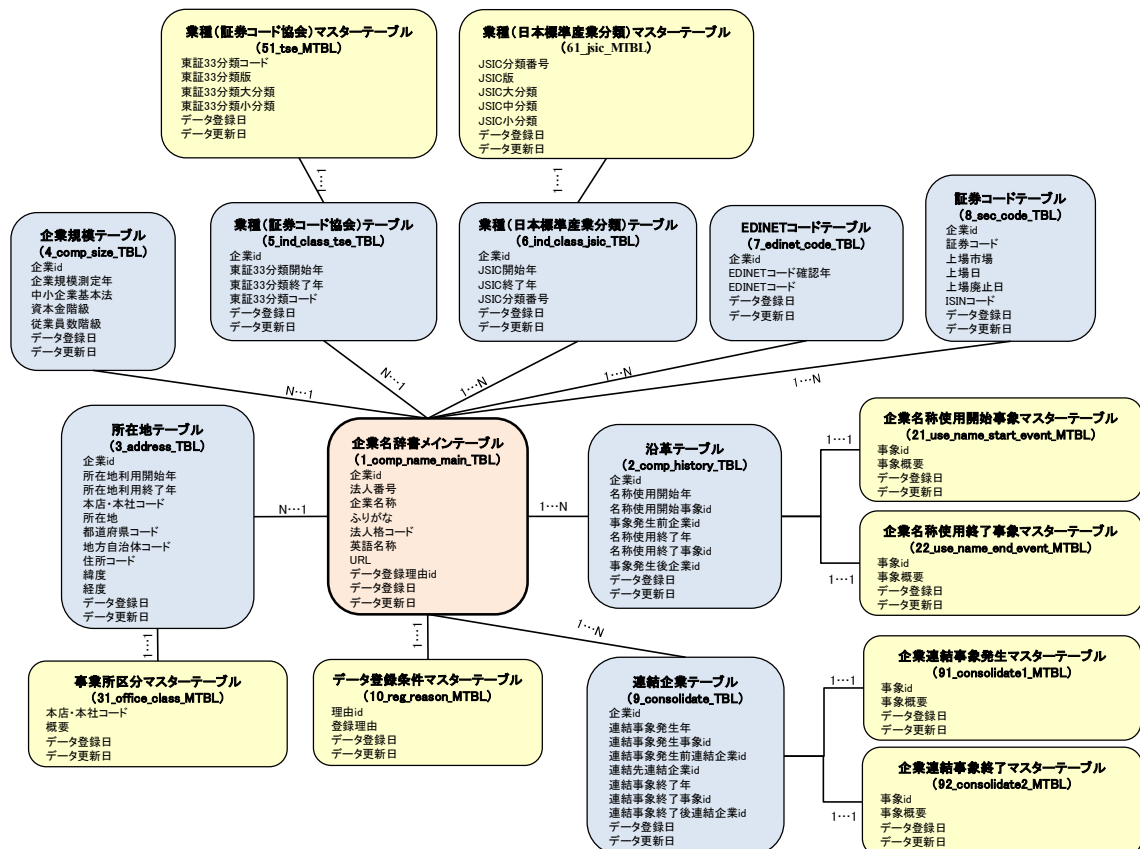
4.2 データベースの構築

企業名辞書を構成する各テーブルファイルは、MySQL 等のデータベース管理システムを用いてインポートし、**図表 25** に示すテーブル間のリレーションを定義する。

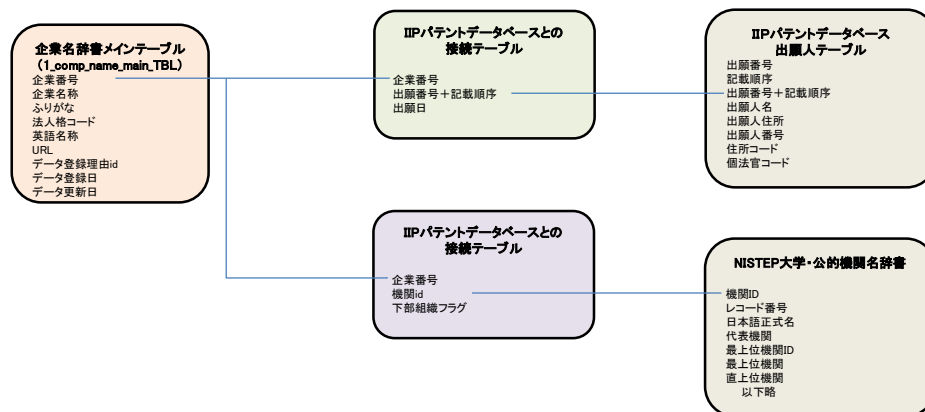
データのインポート方法や構築方法は利用するデータベース管理システムの説明書を参照されたい。

IIP パテントデータベースと接続する場合は、**図表 26** に示すように接続テーブルを介して接続を行う。また、市販の財務データベースやその他企業情報と接続する場合は、法人番号、証券コード、EDINET コードを利用して接続することができる。

図表 25 企業名辞書 E R D



図表 26 企業名辞書と IIP パテントデータベース／NISTEP 大学・公的機関名辞書との接続



5. 企業名辞書の作成と名寄せ

5. 1 企業名辞書の作成

- (1) 企業名辞書に掲載する特許出願実績の多い企業、及びある時期に特許出願数が増加した企業の特定は、IIP パテントデータベースを用いて行っている。
- (2) 企業の特許出願数は、企業名称の変遷、合併、及びパテントデータベースの出願人表記揺れ等を考慮した企業名寄せを行い算出している。
- (3) 企業名称の表記は、法人番号登録名称、有価証券報告書、又は企業がインターネット等で用いられている表記としており、登記上の表記とは異なる場合もある。
- (4) 企業名称の読みに関し、「日本」の読みは、下記URL掲載の「日本（にほん・にっぽん）の読み方」に掲載の情報を参考とした。「にっぽん」として掲載のない企業名称は、原則、「にほん」としている。

<http://hiramatu-hifuka.com/onyak/nippon.html>

- (5) 緯度・経度

所在地の緯度・経度の算出には、東京大学空間情報科学研究センターが下記URLにて利用提供する「CSV アドレスマッチングサービス」を利用している。

<http://newspat.csis.u-tokyo.ac.jp/geocode/>

- (6) 企業の産業分類（日本標準産業分類）は、@nifty ビジネスの企業検索、Wikipediaの日本の企業一覧、当該企業のホームページなど各種のインターネット情報から取得した情報を参考として、総務省統計局の日本標準産業分類五十音索引表（分類検索を簡便化するための細分類各項目の具体的業種が例示された表）と引き当てした産業分類としている。

5. 2 企業名の変遷を考慮した名寄せ

(1) IIP パテントデータベースの特許出願人データは、整理標準化データ⁴を使用しているが、同一出願人に対して異なる表記（表記揺れ）が存在し⁵、同一出願人単位の集約を行う場合には表記揺れをクレンジングした上で「名寄せ」する作業が必要となる。また、整理標準化データでは出願人固有の ID として出願人番号が付与されているものの、年代を遡るほど出願人番号が付与されていない例も多数存在する。このため、出願人を適切に同定するために以下のような手順で名寄せ作業を実施した。

①企業名の修正と標準化

特許出願人名について、目視による修正を行った。目視による修正は、類似表記が多数ある場合は、同一年における出願数の多い表記を正として、その表記に統一した。次に、「株式会社」「有限会社」などの法人格を表す表記を標準コード化し、企業名の表記法として<企業名+法人格コード>の形式を本整備の標準形式とした。

②所在地表記の修正と標準化

多様な表記のゆれが混在する IIP パテントデータベースの出願人所在地表記を対象に、漢数字、ローマ数字、番地町名形式等の表記の標準化を行った。次に市町村合併、町名変更などの情報を別の情報源から収集し、これを用いて、過去の所在地情報も最新のものに統一した。その上で、国土交通省が定義する 12 桁の住所コード（街区レベル位置情報参照情報）を各出願人住所（所在地）に割り当てた。以上の作業により、所在地情報の表記法として<住所コード（12 桁）>形式を本整備の標準形式とした。

③企業名+法人格コード+住所コードの組み合わせによる企業の名寄せ

上記作業で修正した企業名、法人格コードおよび住所コードを利用し、3 項目が一致するレコードを同一企業であるとみなし、名寄せを行っている。加えて、近年の出願分では出願人番号を加えて名寄せを行っている。

④名寄せ結果の評価と修正

名寄せ結果として、異企業を同一企業と見なすエラーは見当たらない。しかし、同一企業を異企業と見なしてしまうエラーが一部発見された。長期間のデータには、所在地表記に番地などの僅かな違いが存在し、12 桁コードでは細密過ぎることが判明した。結論として、5 桁の市区町村レベルのコード⁶で精度高く名寄せできることが判明し、これによる名寄せを実施した。

また、長期間に渡るデータであるため、多くの企業がその間に所在地移転している。このため、同一企業でありながら寄せができない事例も多々生じたため、

⁴ 特許情報を、XML や SGML といった標準的なデータの記述方法を用いて記述することで整理したデータである。

⁵ 以下の文献に詳述されている。

Onishi, K., Y. Nishimura, N. Tsukada, I. Yamauchi, T. Shimbo, M. Kani and K. Nakamura (2012) "Standardization and Accuracy of Japanese Patent Applicant Names", IIPR Working Paper No. 2012-001.

⁶ JIS 市区町村コードと同じコード

移転状況のチェックを行いつつ人手で名寄せを行った。

- (2) 企業の変遷名称を考慮した同一企業の名寄せは、企業名辞書作成初期の Ver. 2014.1 までは大西宏一郎氏（早稲田大学）、西村陽一郎氏（神奈川大学）らの研究グループにより整備された特許出願人の名寄せに関するデータベース(ONAPPdata)を参考とさせて頂いた。

以上

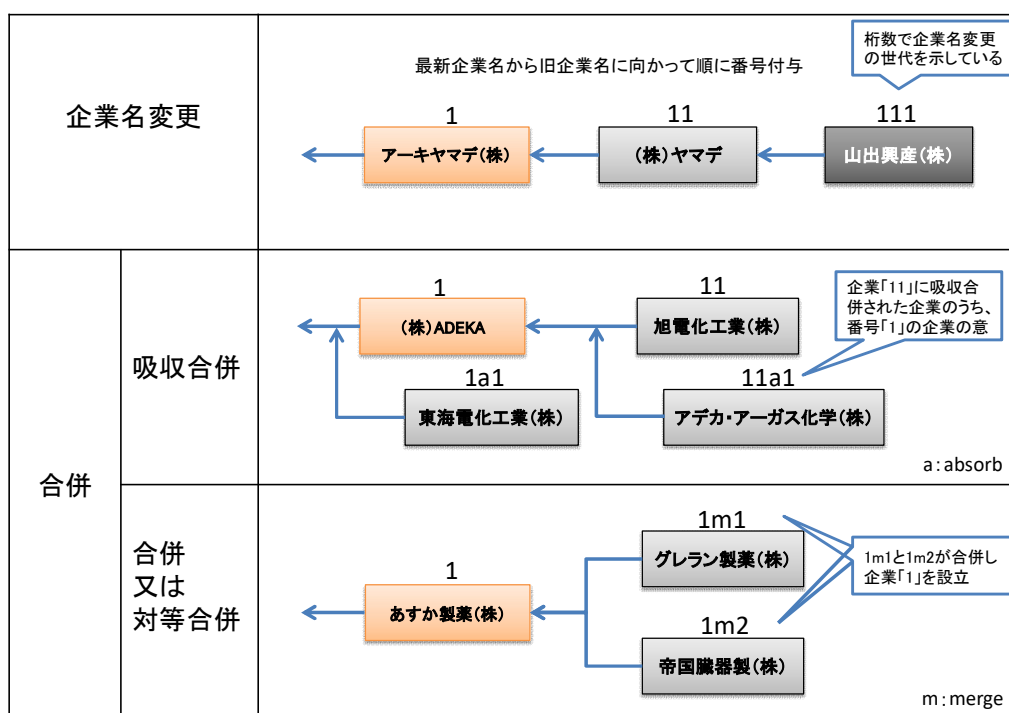
【Appendix】

(1) 沿革 (history) を表現するコード

単一テーブル型の企業名辞書には、企業名称変更時の形態（合併のような形態が変化が伴う場合も含む）を簡易的に示す沿革 (history) フィールドが付加されている。

沿革 (history) に記されたコードの意味は、図表 27 に示す通りである。

図表 27 沿革コードの説明



注: 上記に例示した企業は、企業名辞書の「あ」行掲載企業から例示に適した企業を選択した

【補足説明】

- ・企業名変遷の世代番号は、現在の名称から遡る形態で付与しているため、通常の世代番号とは逆になっている
- ・複数企業の合併において、存続・消滅企業が明確な場合、存続企業が世代番号の継承を行う

(2) 米国特許との接続

米国特許と企業名辞書に掲載する企業との接続は以下のように行っている。

(2) - 1 接続対象米国特許

米国特許の母集団は、2001 年以降に出願された公開特許公報である。

(2) - 2 日米特許の紐づけ

a) 国内特許から米国へパリ優先権を主張して米国へ出願した出願人情報の利用

企業が特許出願を行う場合、自国への特許出願後に、パリ優先権を主張して、海外に権利取得のために特許出願を行うことが多い。米国公報に付与された日本からの優先権主張を抽出することで、国内から米国へ移行された特許を判別できる。優先権主張番号は、国内出願番号であるので、この二つを日米特許の紐づけキーとする。

b) 国際出願 (PCT 出願) から米国移行した特許情報の利用

近年、多数の外国へ特許出願することが多く、その際には国際出願が利用される。そ

ここで、企業が国際出願した後、米国へ移行した特許情報を JP の国際優先権情報が付与された米国公報を抽出することで、日米特許の紐づけキーとする。

c) 日米特許の紐づけ

2001 年以降の米国公開公報の XML データ 4,714,118 件から、米国出願番号、公開番号、パリ優先権主張番号、国際優先権主張番号、米国出願人名 (applicant 又は assignee) と、発明者名を取得する。

これらの情報と、(2) - 2 a) 及び b) で取得した優先権番号すなわち国内出願番号をキーとして紐づけする。

紐づけの結果、529,316 組の日米特許のペアが生成された。

(2) - 3 企業名辞書掲載企業と米国特許出願人 (applicant 又は assignee) の紐づけ

529,316 組の日米特許出願のペアから、日米ともに単願である 354,698 組及び日米ともに 2 出願人である 5,216 組について企業名辞書掲載企業について紐づけを行った。なお、2 出願人の共願のペアにおいては、出願人名の順番が入れ替わっているものがあり、入れ替え修正を行っている。

(2) - 4 接続テーブルの作成

(2) - 3 の結果を利用して、企業名辞書と米国特許との接続テーブルを作成した、なお、米国特許には米国出願番号及び米国公開番号を付与した。本来、米国出願番号のみで十分なはずであるが、米国出願番号のシリーズコードには不備があることが既知であり、公開番号の付与も行っている。

(3) 企業名辞書を利用するための参考資料

企業名辞書を利用するための参考資料として、下記を示す。

- ・中山保夫・富澤宏之 (2018) 「客観的根拠 (エビデンス) に基づく政策のためのデータ・情報基盤 (第一回) ~ NISTEP 企業名辞書 ~」, 『STI Horizon』2018 夏号 (Vol. 4 No. 2), p. 47-53, 文部科学省科学技術・学術政策研究所 (NISTEP).

以上

レポート

客観的根拠（エビデンス）に基づく政策のための データ・情報基盤（第一回） ～ NISTEP 企業名辞書～

第2研究グループ 客員研究官 中山 保夫、総括主任研究官 富澤 宏之

概要

科学技術・学術政策研究所（NISTEP）では、文部科学省の「科学技術イノベーション政策における『政策のための科学』推進事業（SciREX）」の一環として、政策形成を支えるエビデンスの充実のためのデータ・情報基盤の構築を進めている。その成果として、「NISTEP データ・情報基盤」の公開を行っており、様々な研究者等による活用も進んでいるが、潜在的利用者の掘り起こしの余地は大きいと考えられる。ここでは、三回にわたり、「NISTEP データ・情報基盤」の具体的な利用方法について紹介する。第一回は、産業における研究開発・イノベーションに関するデータ基盤の中核となる「NISTEP 企業名辞書」を取り上げ、紹介する。

キーワード：科学技術イノベーション，政策のための科学，企業名辞書，名寄せ

1. 客観的根拠（エビデンス）に基づく政策の実現に向けて：NISTEP データ・情報基盤

文部科学省が平成 23 年度（2011 年度）より実施している「科学技術イノベーション政策における『政策のための科学』推進事業（SciREX）」は、“客観的根拠（エビデンス）に基づく政策”の実現に向けて、政策研究の推進や専門人材の育成等を一体的に行う取組である。

その一環として、科学技術・学術政策研究所（NISTEP）は、エビデンスの充実や政策研究の高度化のための「NISTEP データ・情報基盤」の構築・公開を進めている。STI Horizon 誌では、本稿を含めて三回にわたり、「NISTEP データ・情報基盤」の主要なコンテンツを取り上げ、具体的な利用方法について紹介する。第一回に当たる本稿では産業における研究開発・イノベーションに関するデータを分析する際の基礎となる「NISTEP 企業名辞書」を紹介する。今後、第二回では大学・公的機関の研究開発に関するデータを分析する際の基礎となる「NISTEP 大学・公的機関名辞書」、第三回では、科学技術に関する施策

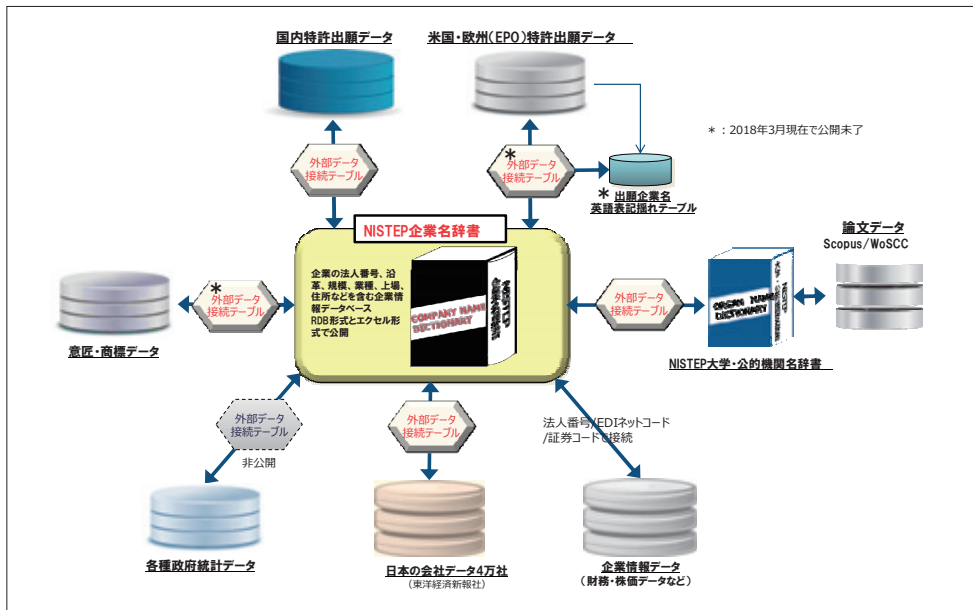
や「NISTEP デルファイ調査」、 「NISTEP 定点調査」に関する各種検索システムを取り上げて紹介する。

2. 産業における研究開発・イノベーションに関するデータ・情報基盤

NISTEP は、産業セクターを対象とした研究開発とイノベーションの実態、イノベーションプロセスなどの分析・研究を行うために、国内営利企業（以下、「企業」と呼ぶ）の知的財産、研究開発投資、事業展開等のデータを体系的に関係付けできる「産業における研究開発・イノベーションに関するデータ・情報基盤」の構築を行っている。

この「産業における研究開発・イノベーションに関するデータ・情報基盤」は、図表 1 に示すように、企業情報検索の索引ともいえる「NISTEP 企業名辞書」（以下、企業名辞書と略す）を核として、企業の研究開発に関するインプット、アウトプット及びアウトカムの各データをリンクし、企業レベルでイノベーションの分析・研究を可能としている。

図表 1 「産業における研究開発・イノベーションに関するデータ・情報基盤」の全体構成図



2-1 名寄せ問題の解決

産業セクターの研究開発やイノベーションに関するデータを対象とした分析・研究を行う場合、企業名の名寄せという問題が立ちはだかる。この問題には二つの側面がある。一つは多くの企業データには誤記も含んだ企業名表記の揺らぎが存在し、同一企業の判断をする上で障害となるという側面、もう一つは統合、合併又は事業再編等による組織形態の変化や名称の変更、場合によっては清算・倒産などの変遷を考慮して企業データを取扱いつつ必要があるという側面である。

前者の例として、特許データの出願人に記載された企業名の表記がある。特に、電子公報発行以前の出願人表記には、片仮名、新字旧字、異称、OCRの精度に起因する誤字等が数多く存在し、特許分析の大きな障害となっていた。このため、膨大な数の企業出願人の表記にクレンジングを施し、その企業名と住所コード及び出願人コード等の判別条件を組み合わせた機械的判定による企業の名寄せに加えて、機械的判定では困難な企業の名寄せを目視作業により行い、この問題の解決を図っている。

後者については、後述する企業名辞書において、現行企業、変遷企業及び被合併企業ごとに当該企業の外部データと接続でき、それら企業を個別に、又はグループ化して関係データの取扱いができる構造とすることで問題解決している。

これにより、本データ・情報基盤の利用者は、企

業名辞書と図表 1 で想定する外部データを接続して利用する限り、名寄せという難しい問題から解放される。

2-2 企業名辞書

企業名辞書は、企業の規模、業種、所在地、上場及び親会社、子会社などの企業の基本情報の提供、並びに特許データ・財務データなどの外部データを接続するハブとしての役割を担う。

企業名辞書では、以下の条件のいずれかに該当

することを企業掲載の基準としており、最新の企業名辞書^{注1}では企業数 10,064 社、それらの変遷企業及び被合併企業も含む全企業レコード数は 20,703 を数える。なお、下記①及び③に該当する企業の抽出には、前述の企業名の名寄せによる成果を使用している。

- ①特許出願数累積 100 件以上 (1970 年以降)
- ②株式上場企業 (2012 ~ 2017 年に上場している企業)
- ③特許出願数の伸び率大 (3, 5, 7 年の各期間で評価)^{注2}
- ④「NISTEP 大学・公的機関名辞書」掲載企業 (論文執筆者所属企業)

2-3 外部データとの接続テーブル

外部データとの接続テーブル (以下、「接続テーブル」と呼ぶ) は、企業名辞書に登録した企業と特許データ等の外部データに収められた当該企業 (特許の場合は出願企業) との接続を実現するテーブルであり、データ間の架け橋の役割を持つ。また、接続テーブルは、外部データの企業名称の表記揺れを吸収する役割も併せ持つ。例えば、前記の特許データの出願人表記に揺れがあるものの同一企業と判定された企業は、接続テーブルにより企業名辞書の正式企業名称レコードと接続され、表記揺れの吸収が図られている。

注 1 2018 年第 3 四半期に公開予定の企業名辞書 Ver2018_2 における掲載企業数である。本稿執筆時点において精査中の段階にあり、公開までに若干変動する可能性がある。

注 2 伸び率の評価方法は、以下の URL に掲載する「企業名辞書利用マニュアル」を参照されたい。
<http://www.nistep.go.jp/research/scisip/rd-and-innovation-on-industry>

3. 企業名辞書の利用形態

イノベーション分析・研究における企業名辞書の利用形態は大きく二つに分けられる。

図表 2 に示すように、一つは企業の様々な情報を取得する辞書としての利用であり、もう一つは、企業名辞書の特徴である外部データ接続の容易性を活用し、特許、意匠、商標、論文、財務、株価などの外部データと企業名辞書の企業情報とを連携させ、より高度な分析を行うものである。以下に幾つかの利用方法について説明する。

なお、企業名辞書は、企業情報をパネルデータとして保有することや、データの保守性などの観点からリレーショナルデータベース（以下、「RDB」と呼ぶ）として作成し公開している。しかし、利用者全て

が RDB の操作に慣れている訳ではなく、機能が削られても一つのテーブルで企業情報を見通したいというニーズもあるため、単一表形式（Excel 形式）でも提供している。

3-1 企業情報を知る辞書として利用

(1) 企業の変遷

企業名辞書はレコード単位で現行、過去及び被合併法人の企業名称とそれらと関係付ける沿革データを保有しており、容易に企業の変遷を知ることができる。

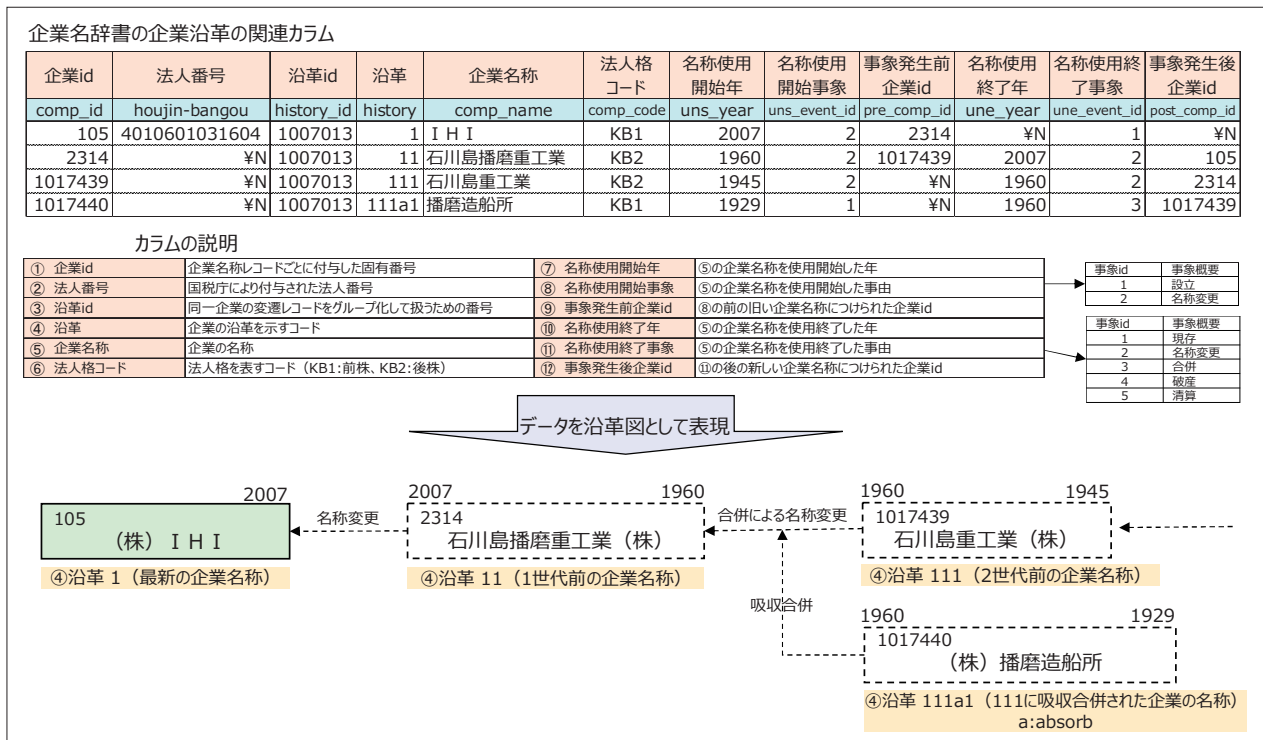
図表 3 は（株）IHI を例に企業名辞書の沿革データを取り出したものである。例えば、データテーブルのカラム'企業 id'^{注3}の値が 105 である（株）IHI のレコードは、「（株）IHI という名称は石川島播磨重工（株）〔'事象発生前企業 id' =2314〕が名称変更〔'名称使用開始事象' =2〕し、2007 年〔'名称使用開始年' =2007〕から使用した名称であり、その名称で企業活動中〔'名称使用終了事象' =1〕である」ことを意味する。

現行、過去及び被合併法人の複数レコードは同一番号

図表 2 企業名辞書の利用形態

(1) 企業名辞書を単独で利用	①法人番号・EDI ネットコード・証券コードなど企業を一意に特定するコード ②企業の変遷名称、合併等の沿革情報 ③住所・住所コード・緯度経度などの所在地情報 ④企業規模・業種などの企業属性情報 ⑤企業の親子関係 ⑥上場企業・上場日・上場廃止日など株式上場情報 ⑦1970年以降特許出願累積数100件以上の企業 など
(2) 企業名辞書と他のデータと連結して利用	①特許出願データと接続して企業の特許出願状況を分析する ②意匠・商標登録データと接続して企業の意匠・商標登録状況を分析する ③財務諸表・株価データ等の企業情報と接続して企業分析する ④論文データと接続して企業の論文算出状況を分析する など

図表 3 企業の沿革を知る



注 3 シングルクォーテーションで囲った文字列はデータテーブルのカラム名を示す。

の'沿革id'でグループ化されており、図表3の'沿革id'=1007013である4レコードを沿革図として表現すると、同図表の下図ようになる。

また、'沿革'の値である沿革コードは変遷を簡易的にコードとして表現したものであり、その意味は図表4を参照されたい。

(2) 企業の規模・業種

図表5に示すように、企業名辞書は企業の規模や業種のデータも保有しており、目的とする企業の情報取得や分析におけるカテゴリーデータとして利用できる。

企業の規模は、中小企業基本法の定義に基づく業種・資本金・従業員数を考慮した分類、資本金階級による分類及び従業員数階級による分類の3種類のデータを保有している。また、業種は、企業の主業について、日本標準産業分類に基づいた大・中・小の3分類に加えて、上場企業については証券コード協議会による業種分類を保有している。

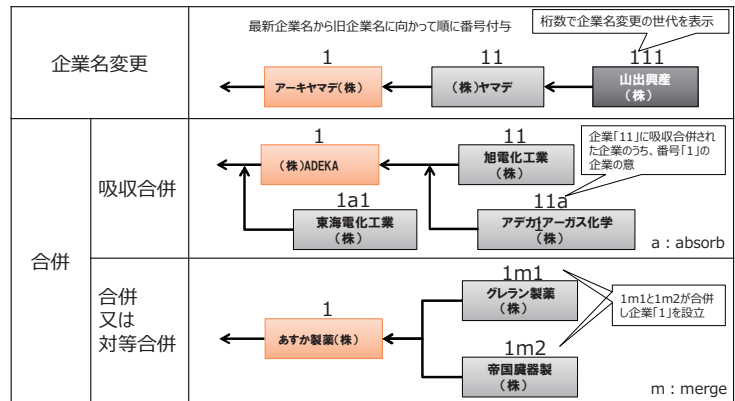
企業の規模や主業は企業活動とともに変化する場合もあり、RDB形式の企業名辞書では調査日ごとにデータをパネル化して保有できるよう考慮している。簡易版の単一表形式では最新調査日のデータが表示される。

(3) 上場企業

企業名辞書には2012年時点で株式上場していた全企業と、以降、企業名辞書の改訂ごとに新規又は再上場した企業を追加掲載している。その間、上場を廃止した企業もあることから、RDB形式の企業名辞書では証券コード(4桁の銘柄コード)、上場市場、上場日、上場廃止日及び証券コード協議会による業種分類の各情報を提供し、ある時点における上場企業を絞り込んで取扱いすることができる。

図表6はExcelの単一表形式の企業名辞書にオートフィルタ機能を使い、現行活動を行っている企業

図表4 沿革コードの説明



注:

- ・企業名変遷の世代番号は、現在の名称から遡る形態で付与しているため、通常の世代番号とは逆になっている
- ・複数企業の合併において、存続・消滅企業が明確な場合、存続企業が世代番号の継承を行う

図表5 企業の規模・業種を知る

企業名辞書の企業規模・業種の関連コラム										
企業id	法人番号	企業名称	法人格コード	企業規模	資本金階級	従業員数階級	日本標準産業分類コード	日本標準産業分類(大分類)	日本標準産業分類(中分類)	日本標準産業分類(小分類)
comp_id	houjin-bangou	comp_name	comp_code	comp_size_law	comp_size_cap	comp_size_emp	jsic_code	jsic_l	jsic_m	jsic_s
1	5120001158218	パナソニック	KB2	大企業	10億円以上	5,000人以上	jsic293	製造業	電気機械器具製造業	民生用電気機械器具製造業

セル値の選択肢		
大企業	100万円未満	5人未満
中小企業	100万円以上	5~29人
小規模企業	1000万円以上	30~99人
	2000万円以上	100~299人
	5000万円以上	300~999人
	1億円以上	1,000~4,999人
	10億円以上	5,000人以上

総務省の日本標準産業分類を記した下記URLを参照
http://www.soumu.go.jp/toukei_toukatsu/index/seido/sangyo/02toukatsu01_03000044.html#

図表6 最新の上場企業を知る

沿革id	沿革	企業名称	法人格コード	名称使用終了	証券コード	証券コード(大)	証券コード(中)	証券コード(小)	企業規模	資本金階級	従業員数階級	日本標準業種	日
you	history_id	history	comp_name	comp_cc	une_event	sec_code	class	s_class	comp_si	comp_si	comp_si	jsic_code	jsic
17640	8002467	1 メディカル・データ・ビジョン	KB2		1	3902	運輸・情報通情報	通信業	大企業	1億円以上	100~299人	jsic391	情報
6966	8000679	1 K Hネオケム	KB2		1	4189	製造業	化学	大企業	10億円以上	300~999人	jsic163	製
1392	8000448	1 オプティム	KB1		1	3694	運輸・情報通情報	通信業	大企業	1億円以上	100~299人	jsic391	情報
5907	8000359	1 エムケイシステム	KB1		1	3910	運輸・情報通情報	通信業	中小企業	1億円以上	30~99人	jsic391	情報
1476	8000277	1 エクストリーム	KB1		1	6033	サービス業	サービス業	大企業	1億円以上	100~299人	jsic391	情報
1663	1009997	1 ヘルーナ	KB1		1	9997	商業	小売業	大企業	10億円以上	300~999人	jsic611	卸
5153	1009996	1 サトー商会	KB1		1	9996	商業	卸売業	大企業	10億円以上	300~999人	jsic522	卸
6937	1009995	1 ルネサスイーストン	KB1		1	9995	商業	卸売業	大企業	10億円以上	300~999人	jsic541	卸
4536	1009994	1 やまや	KB1		1	9994	商業	小売業	大企業	10億円以上	300~999人	jsic585	卸
2171	1009993	1 ヤマザワ	KB1		1	9993	商業	小売業	大企業	10億円以上	300~999人	jsic581	卸
4707	1009992	1 理研グリーン	KB1		1	9992	商業	卸売業	大企業	10億円以上	300~999人	jsic532	卸
		1 マックス	KB1		1	9991	商業	卸売業	大企業	10億円以上	300~999人	jsic532	卸

（‘名称使用終了事象’=1）と証券コードの記載あり（‘証券コード’ <> ¥N）を条件として最新の上場企業^{注4}を抽出した例である。

(4) 企業の所在地情報の利用

NISTEP では、国立大学の研究者による発明に基づく特許出願の状況を網羅的に調査した結果を報告¹⁾している。

図表7は、その報告から、名古屋大学を例とした特許出願に至る共同発明を行った特許出願累積数（2004～2012年度）の上位企業の情報を使い、その所在地を地図上に表示したもので、地域内の産学連携状況についても直感的に知ることができる。

図表7では、企業名辞書の企業所在地の緯度・経度情報を利用して描画^{注5}しているが、企業名辞書ではこのほかにも、企業の所在地情報として、住所、住所

コード（都道府県レベル（2桁）、市区町村レベル（6桁）、国土交通省コード（12桁））が提供される。

なお、企業の所在地は、原則本社の住所としているが、本店、事業所等の住所も提供される。また、移転により所在地が変わることも数多くあり、所在地を企業名寄せにおける同一企業判定の要素とするためには複数の所在地情報を必要とする。このため、RDB形式の企業名辞書では所在地情報もパネル化できる構造としている。

(5) 企業の変遷に沿った特許出願状況の再現

外部データである特許データを用いなくても、企業名辞書と国内特許出願データ対応の接続テーブルのみで企業の変遷に沿った特許出願状況が再現できる。

接続テーブルは、‘企業id’、‘出願番号_出願人順序’、‘出願日’の3つのカラムを有し、‘企業id’と‘出願番号_出願人順序’は

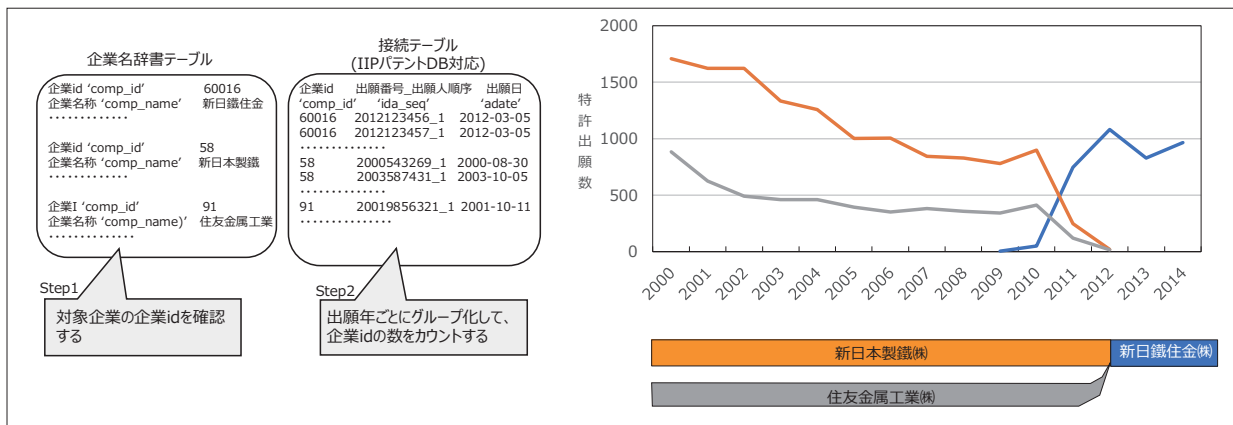
それぞれ企業名辞書と特許出願人テーブルとの接続情報となる。特許の出願人は、原則、公表の公開時点の出願人であるため、企業名辞書の変遷に沿った企業名と接続され、出願日を考慮して接続テーブルの‘企業id’を数えることにより企業の変遷に沿った特許出願状況の再現が可能となる。

図表8は新日鐵住金

図表7 国立大学と共同発明を行った企業の所在地（例：名古屋大学）



図表8 企業の変遷に沿った特許出願状況の再現



注4 厳密には、単一表形式に上場廃止日を含めておらず、上場廃止企業が若干数混在する。

注5 埼玉大学教育学部谷謙二研究室で公開する「Google Maps API v3 を使ったポイント地図化」を用いて描画した。
<http://ktgis.net/gcode/lonlatmapping.html>

(株) について、2000 年からの特許出願状況を示したものである。同社は 2012 年に新日本製鐵(株)と住友金属工業(株)が合併し発足した。図表 8 には発足以前から新日鐵住金(株)名義の出願が見られるが、それらは設立時点で公開未了であった出願について、遡って出願人の名義変更が行われた結果と考えられる。

3-2 外部データと接続して利用

企業名辞書と外部データとの接続は接続テーブルを経由して、又は、市販の財務データやその他の企業データに法人番号、EDINET コード又は証券コードが付されている場合は企業名辞書と直接接続することができる。

法人番号は、国税庁により法人ごとにユニークな番号(13桁)が指定され、個人番号(マイナンバー)と異なり原則公表されるので自由に利用できる。このため、同一企業のデータを接続して利用する際の標準番号として今後利用範囲が広がるであろうが、番号法施行日(平成 27 年 10 月 5 日)以前に消滅した企業は、当然、番号指定はなく、過去に遡った企業データを取り扱う場合はそれだけで十分とはならず他の方法が必要になる。

企業名辞書は、外部データと接続して両者のデータを分析に利用するだけでなく、外部データ接続のハブとして適用し、特許データ、財務データ、意匠・商標データなどを組み合わせて分析することもできる。

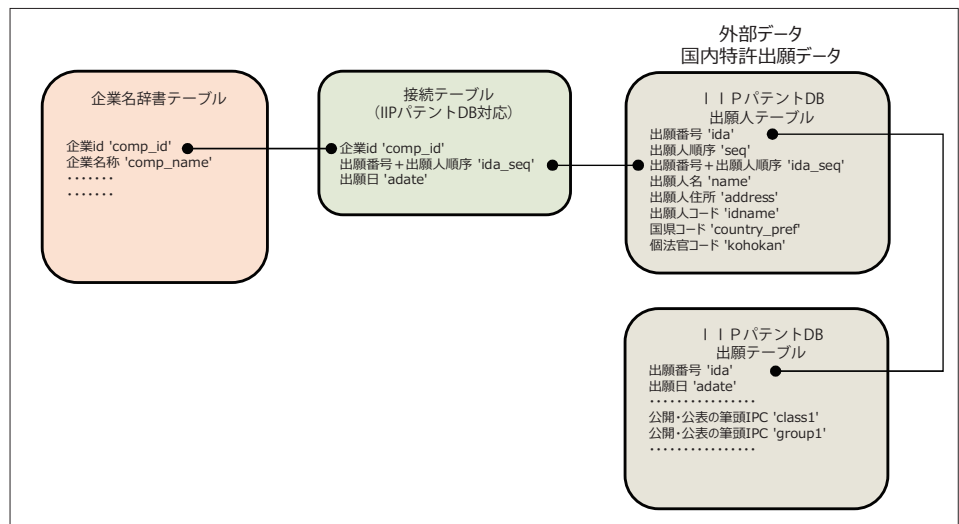
ここでは、国内特許出願データである IIP パテント DB^{注 6}を外部データとして企業名辞書と接続し分析を行った事例を以下に示

す。IIP パテント DB との接続は、図表 9 に示すように専用の接続テーブルを介して、IIP パテント DB の出願人テーブル(applicant table)と接続する。大規模データである特許データを前処理なしに直接分析に供することは本来容易な作業ではないが、企業名辞書が介在することで極めて容易に分析が行える。

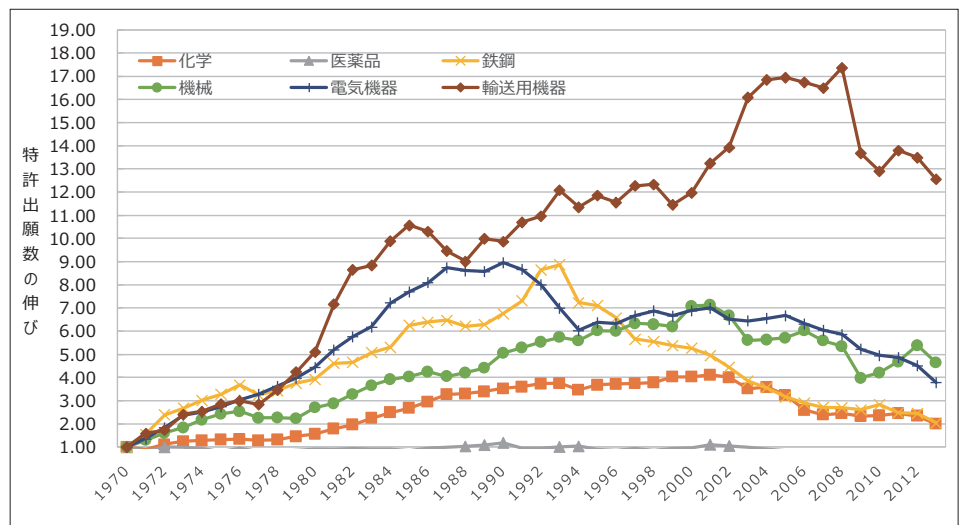
(1) 分析事例：業種別の特許出願数の伸び

図表 10 は、3-1(3)で例示した最新の株式上場企業を対象に、証券コード協会の業種分類(小分類)をカテゴリーデータとして、1970 年の業種ごとの特許出願数を基準に 6 業種について年ごとの出願数の伸びを表示したものである。

図表 9 企業名辞書と IIP パテント DB との接続方法



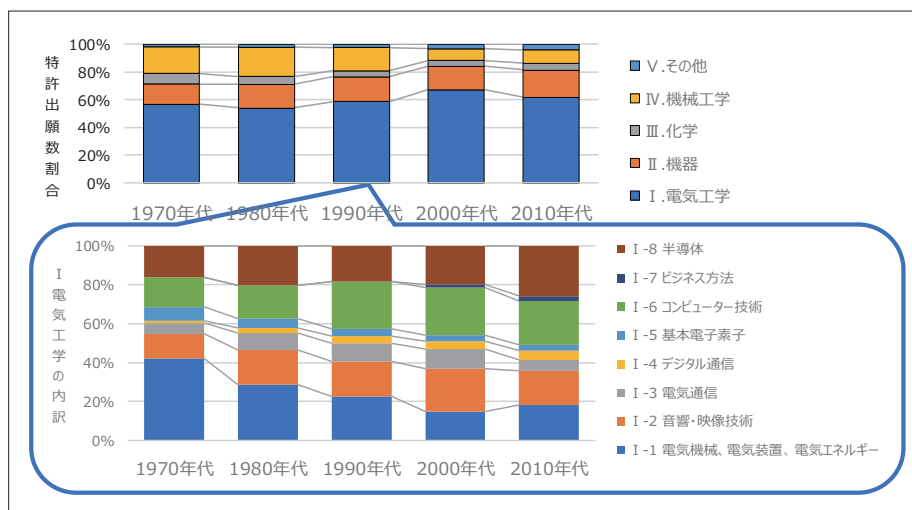
図表 10 業種別の特許出願数の伸び



注：図表 10 は 6 業種について 1970 年の特許出願数を 1 と表示した

注 6 (一財) 知的財産研究所より公開される特許庁の整理標準化データをもとに特許統計分析用に開発されたデータベースである。本稿執筆時点の IIP パテントデータベースの最新版は 2017 年版であり、同研究所のホームページより入手できる。
(<http://www.iip.or.jp/>)

図表 11 ある電気機器メーカーの特許出願技術領域の変化



こうした図表は、例えば景気動向、株価等の経済指標と組み合わせて考察するなど、新たな知見を取得するための材料としても使える。

(2) 分析事例：発明の技術領域の変化

図表 11 は、ある電気機器メーカーが特許出願した発明について、年代ごとに技術領域がどのように変化したかを示している。ここでは、図表 9 のように、出願人テーブルに加えて、発明の技術分類である IPC (International Patent Classification: 国際特許分類) データが記載された出願テーブル (ap table) を接続して利用している。図表 11 下図注⁷ は、I 電気工学分野の細目を示しているが、いわゆる、重電領域から半導体、計算機領域へと比重が移っている様子がうかがえる。

4. おわりに

企業名辞書と接続テーブルは、以下の NISTEP のホームページにて公開しており、記載する条件で使用する限り、誰でも利用することができる。

<http://www.nistep.go.jp/research/scisip/rd-and-innovation-on-industry>

企業情報は常に変化しており、企業名辞書が陳腐化しないよう情報を収集し、一定の品質を担保するには多くの手間を必要とする。外部データも特許データのように年間数十万件のデータが追加されるものもあり、逐次接続情報を追加することが必要になる。企業名辞書の維持に行うべきことはたくさんあり、そのフォローは大変ではあるが、科学技術イノベーション政策の形成に企業名辞書が活用され、貢献することができれば幸甚である。

注 7 IPC を WIPO (World Intellectual Property Organization: 世界知的所有権機関) の「IPC - Technology Concordance Table3」で変換した技術分野で描画している。

参考文献

- 1) 科学技術・学術政策研究所, 『国立大学の研究者の発明に基づいた特許出願の網羅的調査』, 調査資料 -266, 科学技術・学術政策研究所, 2017 年 12 月.
DOI: <http://doi.org/10.15108/rm266>