

大学・公的機関名英語表記ゆれテーブル(Ver.2016.1)利用の手引き

2016年10月31日

文部科学省科学技術・学術政策研究所

1. はじめに

研究論文等のデータベースの利用に際して、機関名で検索したり、機関別の集計や分析を行ったりすることがよくあります。そのときの厄介な問題の一つは、機関名の表記が統一されておらず、いろいろな「表記のゆれ」が見られることです。英語のデータベースで、たとえば東京農工大学の正式英語名は Tokyo University of Agriculture and Technology ですが、これが Tokyo Noko University、Tokyo Agriculture and Technology University などと表記されたり、“University”が”Univ”、“Agriculture and Technology”が”A&T”などと略記されたりします。この問題に対処するため、Web of Science Core Collection (WoSCC)や Scopus などのデータベースでは、各機関に固有の識別番号を与える、この識別番号による所属機関の一括検索機能を設ける等、機関検索の精度向上が図られています。しかし、あるテーマに関する検索結果をダウンロードしてその中の機関構成を分析したいときなどは、これらの機能を使うことができません。また、機関の下部組織等についてまで識別がなされているわけではありません。

科学技術・学術政策研究所(NISTEP)が「科学技術イノベーションにおける“政策のための科学”推進事業」の一環として進めている「公的研究機関に関するデータ整備」では、その整備事業の中で、データベースで使われている機関表記のゆれを調査・分析しています。2015年10月に、「NISTEP 大学・公的機関名辞書」(以下「機関名辞書」と呼びます)に含まれている機関の英語表記(正式名の他、通称、略称等の別名を含む)と、WoSCC データベースに現れる主な機関名英語表記のデータを統合し、「大学・公的機関名英語表記ゆれテーブル(Ver.2015.1)」を公開しました(公開に当たり、WoSCC の提供元である Clarivate Analytics(旧トムソン・ロイター IP&Science)の了解を得ております)。

今回、この表記ゆれテーブルのうち機関名辞書由来のデータを、同辞書の更新(Ver.2015.1 → Ver.2016.1)に合わせて更新し、「大学・公的機関名英語表記ゆれテーブル(Ver.2016.1)」(以下、混同の恐れがなければ単に「テーブル」と略記)として公開いたしました。以下をお読みの上、機関名辞書(Ver.2016.1)と合わせて、我が国の研究機関に関する情報分析に活用していただくことを希望しております。

なお、このサイトでは、Scopus データベースに現れる表記ゆれの調査結果をまとめた「大学・公的機関名英語表記ゆれテーブル(Scopus 版)(Ver.2013.1)」を Elsevier 社の了解を得て公開しています。この Scopus 版テーブルは、大学と公的機関のみを対象にしていること、機関の表記ゆれの中に一部その下部組織に対する表記が含まれていることなど、データ収録の基準にやや異なる点があるため、本テーブルと統合せず別途に公開しています。これらの点を修正した後、すべてを統合した機関名英語表記ゆれテーブルとする予定です。

※このテーブルの利用について



大学・公的機関名英語表記ゆれテーブル(Ver.2016.1)の利用について

ては、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」を適用します。すなわち、原作者のクレジットを表示し、改変した場合にはこのテーブルと元の作品と同じCCライセンス(表示-継承)で公開することを守れば、営利目的での二次利用も許可されます。表示するクレジットは次のようになります。

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: 大学・公的機関名英語表記ゆれテーブル(Ver.2016.1)

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

CCライセンスと、このライセンスのコモンズ証、リーガルコードについては、

<http://creativecommons.jp/licenses/> をご覧下さい。

2. 収録対象機関

このテーブルに収録対象としている機関は、原則として、機関名辞書に収録されている機関と同じです。すなわち、研究活動を行っている我が国の機関が対象です。大学等(短期大学、高等専門学校、大学共同利用機関を含む)、国の機関、特殊法人・独立行政法人(国立研究開発法人を含む。以下同じ)の他、地方自治体の機関、非営利法人、民間企業も含まれます。以下では、これらの独立した機関を「代表機関」、代表機関に属する組織を「下部組織」と呼びます。単に「機関」というときは、代表機関と下部組織の両方を意味します。このテーブルには、大学、大学共同利用機関、国の機関、特殊法人・独立行政法人の主要な下部組織も含まれています。

機関名辞書の区分に従い、以下の機関は下部組織ではなく代表機関としています。

- (a) 大学の一部としての短期大学部、高等専門学校
- (b) 国立高等専門学校: これらは独立行政法人国立高等専門学校機構の下部組織ですが、ここでは代表機関としています。
- (c) 国立試験・研究機関: 但し、試験・研究機関に属しない国の機関(気象庁地磁気観測所等)は属する省庁の下部組織としています。
- (d) 地方公共団体の公設試験研究機関等

3. データ源と抽出した機関英語表記データ

1. で述べたように、このテーブルは、もともと機関名辞書に含まれていた機関表記(データ源が「辞書」となっているデータ)と WoSCC 版に含まれていた機関表記(データ源が「WoS」となっているデータ)を合体したものです。ここでは、それぞれのデータ源における機関表記データについて説明します。

3.1 機関名辞書の機関英語表記データ

機関名辞書に含まれている英語の正式機関名称とそれ以外の「別名」(通称、略称等)を含みます。正式名称の場合、テーブルの「正式名称」欄に○印が付いています。但し、○印が付いているのは当該機関の Web サイトにより正式名称であると確認された表記なので、○印が付いていなくても正式名称である場合があります。地方公共団体の機関、会社、非営利法人に属する機関の表記にはこのような例が多く含まれます。

別名はいろいろな情報源から採られています。しかし、このテーブルには、機関名辞書中のすべての別名表記を含むものではありません。NISTEP では、機関名辞書を用いてデータベース中の著者所属機関データの同定を行っていますので、その際同定漏れや誤同定を防止するためにいろいろな別名データを収録していますが、これらはこのテーブルには取り入れておらず、機関の略名、通称、旧名などに限っています。詳しくは付録 1 をお読みください。なお、3.2 で述べる WoSCC 版の機関表記データには、ここで除いたタイプの別名表記も多く含まれています。

正式名称か別名かに拘わらず、会社を表す"Co., Ltd."、"Corp"等を除いて、機関の法人格を示す接頭辞または接尾辞は原則として付いていません。大学の下部組織名称には、下部組織名の後に上位の大学名が付けられています（たとえば Graduate School of Medical Sciences, Kyushu University）が、その他の機関では、下部組織名称に代表機関名が含まれているかどうかは不定です。

3.2 WoSCC 版の機関英語表記データ

データ源は、1996-2012 年の期間に WoSCC データベースに採録された論文のうち、日本の機関に所属する著者を含む論文のデータです。これに該当する論文は約 150 万件、その中の日本機関のデータは延べ 275 万件以上になります。

これらの論文における著者所属機関データの機関同定（所属機関表記を機関名辞書の登録機関に対応付けること）によって得られた結果が、このテーブルに含まれるデータです。但し、同定される表記データの種類は膨大な数になるので、このテーブルに記載する機関表記を次の手順で選択しました。

- (1) 同定された機関データが延べ 1,000 以上の機関を選択します。このとき、下部組織はその代表機関に、現存しない機関は最終継承機関に合体して集計します。
- (2) (1)で選択した機関データから、10 回以上出現した表記バリエーションを取り出します¹。
- (3) これらの表記バリエーションを、機関名辞書に登録した機関単位に分けます。代表機関とその下部組織、非現存機関とその継承機関は、それぞれ別の機関単位です。

これにより選択された表記バリエーションは 2,217 となりますが、それらによりカバーされる WoSCC 中の出現頻度は、日本機関データ全体の 68%、このテーブルに存在する機関のデータの 75%になります²。つまり、WoSCC でこれらの表記バリエーションを含めた検索や抽出を行えば、対象の機関の大部分の論文データをカバーすることができます。

¹ 但し、下部組織については、代表機関の直下の下部組織の表記のみを取り出しています。たとえば、大学院研究科名を表す Grad Sch Sci という表記バリエーションは取り出しますが、研究科名とその下の専攻科名が結合した Grad Sch Sci Dept Phys や、専攻科名のみ Dept Phys という表記バリエーションは原則として取り出していません。

² 代表機関のカバー率はこれよりずっと高いのですが、下部組織では表記は極めて多様なためカバー率が低くなります。下部組織の表記ゆれの分析については、『NISTEP 大学・公的機関名辞書の整備とその活用－大学下部組織レベルの研究データ分析に向けて－』, NISTEP NOTE-15, 科学技術・学術政策研究所, 2015 年 10 月 (<http://data.nistep.go.jp/dspace/handle/11035/3085>)を参照してください。

WoSCC の検索やそれを用いたデータ分析にこのテーブルを利用される場合は、付録2をお読みください。

4. テーブルの概要

4.1 機関数と表記バリエーション数

テーブルに記載される機関は13,647(代表機関11,182、下部組織2,465)です。また、表記バリエーションの総数は17,177(代表機関13,265、下部組織3,912)で、15,122が機関名辞書から、2,141がWoSCCからのデータです(両者に含まれている表記が86)。

4.2 データ項目

テーブルには以下のデータ項目が含まれます(それぞれの内容については5.をお読み下さい)。

[A] 機関表記に関するデータ項目

- ・表記バリエーション
- ・正式名称であるかどうか
- ・データ源が機関名辞書、WoSCCのいずれか
- ・WoSCCにおける出現頻度(データ源がWoSCCの場合)

[B] 機関に関するデータ項目

- ・機関ID
- ・機関の日本語正式名称
- ・機関が属するセクター
- ・代表機関か下部組織かの別(下部組織の場合その代表機関を記載)
- ・現存機関か非現存機関かの別(非現存機関の場合その最終継承機関を記載)

5. テーブルの見方

テーブルはExcel形式(.xlsx)で単独のシートから成ります。データ項目(テーブルの各列)は、機関表記に関する項目と機関に関する項目に大別されます。

5.1 各データ項目の説明

[A] 機関表記に関するデータ項目

- (1) 表記バリエーション:機関名辞書またはWoSCCから抽出された英語の機関表記です。詳しくは3.をお読み下さい。
- (2) 機関内番号:同一機関の表記バリエーションに付けられる一連番号です。正式名称がある場合はその番号が1、あとはアルファベット順です。
- (3) 正式名:その表記バリエーションが正式名称の場合“○”が付いています。正式名称については3.1の説明をお読み下さい。
- (4) データ源:その表記バリエーションが採られたデータ源で、次のいずれかです。
 - ・辞書、WoS:機関名辞書とWoSCCの両方に含まれる表記バリエーション

・辞書:機関名辞書のみに含まれる表記バリエーション

・WoS:WoSCCのみに含まれる表記バリエーション

(5) WoS 頻度:データ源が「辞書、WoS」または「WoS」の場合、その表記バリエーションが WoSCC に出現した回数です。詳細は 3.2 の説明をお読み下さい。

[B] 機関に関するデータ項目

以下の項目についての詳細な説明は、このサイトにある「NISTEP 大学・公的機関名辞書(Version 2016.1)利用マニュアル」をご覧ください。

- (1) 機関 ID:機関名辞書で与えられている各機関の識別番号です。番号の付与方法に特別のルールはありません。
- (2) 機関名:機関(代表機関または下部組織)の日本語正式名称です。
- (3) セクター番号及びセクター:機関が属するセクターとそれらに付けられた番号で、次のようになっています³。

	セクター番号	セクター
大学等	1	国立大学
	2	国立短期大学
	3	国立高等専門学校
	4	公立大学
	5	公立短期大学
	6	公立高等専門学校
	7	大学共同利用機関
	12	私立大学
	13	私立短期大学
	14	私立高等専門学校
公的機関	8	国の機関
	9	特殊法人・独立行政法人
その他の機関	10	地方自治体の機関
	15	会社
	16	非営利団体
	17	その他の機関

- (4) 代表機関:“○”が付いている機関は代表機関です。下部組織の場合は空白です。
- (5) 代表機関名:下部組織の場合、それに対する代表機関の名称です。代表機関の場合は「機関名」の名称と同じです。
- (6) 現状:機関が現存する場合は空白、現存しない場合は“No”です。
- (7) 最終継承機関:「現状」が“No”の場合、その継承機関があれば、最終の継承機関を記入しています。

³ セクター11(学校法人)の機関は機関名辞書に英語名称が付いていないので、このテーブルには含まれません。

5.2 レコード(テーブルの行)の単位と並び順

一つのレコード(テーブルの 1 行)は一つの表記バリエーションに対応します。複数の表記バリエーションを持つ機関はその数だけの行数を占め、機関内番号で各行が識別されます。このような機関では、5.1 の[B]に示した機関に関するデータ項目には同じデータが並ぶことになります。やや冗長ですが、自由にソートができるようこのような構成にしています。

レコードは次の順序に配列されています。

- ① セクター番号の順
- ② 各セクターの中で代表機関名の漢字コード順
- ③ 各代表機関の中でまず代表機関(○印)、残りは機関(下部組織)名の漢字コード順
- ④ 各機関内でまず英語正式名称表記(○印)、残りは表記バリエーションのアルファベット順

6. 補足

今回の更新では、WoSCC からの情報は更新されておらず、3.2 で述べたように、1996-2012 年の期間に採録された論文のデータによるものです。また、別途公開している「大学・公的機関名英語表記ゆれテーブル(Scopus 版)(Ver.2013.1)」は、1996-2012 年の期間に Scopus データベースに採録された論文のデータによるものです。これらについては、より最近までの情報を加えて、すべてを統合した大学・公的機関名英語表記ゆれテーブルとすることを考えています。

【大学・公的機関名英語表記ゆれテーブル改訂履歴】

2012/12/18 NISTEP 大学・公的機関名辞書 Ver.2012.1

2013/12/10 大学・公的機関名英語表記ゆれテーブル(WoS版)(Ver.2013.1)

2014/11/14 NISTEP 大学・公的機関名辞書 Ver.2014.1

2014/11/14 大学・公的機関名英語表記ゆれテーブル(WoS版)(Ver.2014.1)

2015/10/15 NISTEP 大学・公的機関名辞書 Ver.2014.1 の英語名称データと大学・公的機関名英語表記ゆれテーブル(WoS版)(Ver.2014.1)を統合し、大学・公的機関名英語表記ゆれテーブル(Ver.2015.1)として公開

2016/10/31 NISTEP 大学・公的機関名英語表記ゆれテーブル(Ver.2016.1)を公開(NISTEP 大学・公的機関名辞書の更新(Ver.2015.1→Ver.2016.1)に伴う更新)

【付録1】機関名辞書収録の別名表記のうちこのテーブルに含むもの

NISTEP 大学・公的機関名辞書に含まれる正式名以外の英語名称は、次のようなタイプに分けることができます。

- I (略称):たとえば、奈良先端科学技術大学院大学に対する NAIST、高エネルギー加速器研究機構に対する KEK など。
- II (通称、旧名等):一般によく使われている別名、機関の旧名等。また、確認できないが正式名ではないかと思える別名も含む。
- III (一部の語の省略):たとえば、正式名 **XX Research Organization** を **XX Organization** と省略した表記、正式名の末尾の", Japan"を省略した表記等。
- IV (一部異なる語):たとえば、国立研究開発法人国立環境研究所の正式名 **National Institute for Environmental Studies** に対する **National Institute for Environmental Sciences** など。Sciences と Science、for と of の違いなども含む。
- V (短縮形):たとえば、国立研究開発法人物質・材料研究機構に対する **Natl Inst Mat Sci** のように語を短縮した表記。
- VI (上位機関名との組み合わせ省略):たとえば、独立行政法人日本貿易振興機構アジア経済研究所の正式名である **Institute of Developing Economies, Japan External Trade Organization** の上位機関部分を省略した **Institute of Developing Economies** のような表記。逆に、下部組織名のみを正式名に上位機関名を付加した表記や、正式名中の上位機関部分と下部組織部分の配列順を入れ換えた表記もある。
- VII (下部組織):機関名辞書に収録されていない下部組織の名称による表記。機関名辞書では、原則として第 2 階層の下部組織(代表機関の直下の組織)までを収録範囲としているが、WoSCC などでは、大学の第 2 階層(学部や研究科)を省略してその下の階層名(学科や専攻など)を示す表記が見られるため、同定の必要上これらを別名としている。たとえば、**Department of Civil Engineering, University of Tokyo** を東京大学工学系研究科の別名にするなど。
- VIII (ミススペル等):単純なスペルミスその他、**Kyushu University** を **Kyusyu University** とする等のローマ字書法の揺れ、単語間のスペースの有無の違い等を含む。

機関名辞書に含まれるこれらの別名表記について、次の基準でこのテーブルに収録しています。

- ・タイプ I とタイプ II はすべて収録。
- ・タイプ III とタイプ IV は個別に検討して重要と考えられるものを収録。
- ・タイプ V, VI, VII, VIII は収録しない。

なお、WoSCC から収録の表記は、このような考えによらず、3.2 に示したとおり、一定以上の出現頻度を持つすべての表記を収録しています。

【付録2】このテーブルのデータを WoSCC の検索に利用する場合の注意

WoSCC の機関検索にこのテーブルに含まれる表記バリエーションを利用すれば、かなり高い再現率を得ることができます⁴。ここでは、その際に注意すべき点について述べます。

A1. 機関名中の語の短縮形表記について

WoSCC では、機関名の表記に次のような短縮形表記法を用いています。このテーブルにおける WoSCC をデータ源とする表記もこの表記法に従っています(但しそれからゆれた表記もあります)

(1) よく出現する語の短縮形表記

代表的な例は次の通りです。

University → Univ

Institute → Inst

National → Natl

Science, Scientific → Sci

Technology, Technological → Technol

Medicine, Medical, Medicinal → Med

And → &

WoSCC の検索で検索項目に所属機関を選ぶと、「短縮形リストを表示」という案内が示されます。このリストから、使用されている短縮形を知ることができます。但し、検索をする場合は原形、短縮形のどちらを用いても構いません(検索語に University、Univ のどちらを用いても同じ結果が得られます)。

(2) 前置詞、冠詞の省略

前置詞の of、for、冠詞の the 等は省略されます。

(3) ハイフン、アポストロフィー等の省略

語中に含まれるハイフン(-)やアポストロフィー(')等は省略されます。ハイフンを省略したとき、その両側を詰める場合とスペースに置き換える場合があります。

このような略記を用いているため、機関名はたとえば次のように表示されます。

The University of Tokyo → Univ Tokyo

National Institute of Advanced Industrial Science and Technology

→ Natl Inst Adv Ind Sci & Technol

The University of Electro-Communications → Univ Electrocommun

Nara Women's University → Nara Womens Univ

A2. 代表機関と下部組織の表記

(1) WoSCC の著者所属機関データフィールド

WoSCC の XML データファイルでは、著者所属機関データを、`affiliation_org`、

⁴ WoSCC の書誌データの利用については、利用機関と Clarivate Analytics(旧トムソン・ロイター IP&Science)が取り交わす利用規約が適用されます。

affiliation_suborg などいくつかのサブフィールドに分割しています。原則として、affiliation_org サブフィールドには代表機関、affiliation_suborg サブフィールドには下部組織の名称が記載されていますので、このテーブルの代表機関と下部組織の表記も、多くはそれに従って抽出されています。この原則に従っていない場合(affiliation_org サブフィールドと affiliation_suborg サブフィールドの間で下部組織名と代表機関名が逆転している場合、affiliation_suborg サブフィールドに代表機関と下部組織の合体した名称が入っている場合など)は、修正して記載しました。

(2) affiliation_org サブフィールドの表記

次のように、代表機関以外を示す語が混入している表記も含まれています。

- ・機関名と下部組織名が合体した表記

[例] Hirosaki Univ Hosp
Kagoshima Univ Museum
Kinki Univ Sch Med

- ・所在地の一部が混入した表記

[例] Kyoto Univ Katsura
JAERI Takasaki

(3) affiliation_suborg サブフィールドの記述

原則として、その代表機関の情報は含まれていません。例えばある大学の工学研究科の場合には Grad Sch Engn となっています。これだけで検索すればあらゆる大学の工学研究科のデータが出てくることとなります。まず、目的とする大学の表記を代表機関テーブルから選択して検索した後、その結果に対して下部組織の表記で絞り込み検索を行うなどの方法をとります。

大学の affiliation_suborg サブフィールド表記には、大学直下の組織(学部や大学院研究科)ではなく、その下の学科や専攻科が記載されていることが多いのですが、このテーブルに含まれるのは、原則として代表機関直下の下部組織です(例外もあります)。

A3. 混同しやすい英語機関名

次のように、異なる機関が同一または類似の英語名称を持つと、それらの名称で検索した場合ノイズ(目的以外の機関の混入)や検索洩れが生じやすくなりますので、注意が必要です。

(1) 同一名称を持つ異なる機関

統合や改組を行った機関が、日本語機関名は変更したのに英語名はそのままという例はよくあります。たとえば、東京都立大学と首都大学東京はどちらも Tokyo Metropolitan University です。また、国立研究所の宇宙科学研究所は独立行政法人宇宙航空研究開発機構の下部組織に移行しましたが、Institute of Space and Astronautical Science の英語名はそのままです。

このような継承関係がない機関が全く同じ英語名を持つことは、大学や公的機関ではほとんどありませんが、会社や非営利団体ではいくつかあります。

(2) 類似の名称を持つ機関

最も注意を要する例として静岡大学と静岡県立大学があります。英語の正式名称はそれぞれ Shizuoka University、University of Shizuoka なので、WoSCC ではそれぞれ Shizuoka

Univ、Univ Shizuokaと表記されることになりましたが、これらに所属する著者が、論文の所属機関に誤って違う大学の名称を記載していることがあります。このような誤記は、データベース作成時にも修正が困難ですので、この両大学では、一方の名称で検索した結果にはある程度のノイズや検索洩れがあります。

他にも、東京農工大学(WoSCCの表記はTokyo Univ Agr & Technol)と東京農業大学(WoSCCの表記はTokyo Univ Agr)なども混同しやすい例です。

(3) 英語名がごく一般的な機関

たとえば、分子科学研究所、厚生労働省国立公衆衛生院(現在は厚生労働省国立保健医療科学院)のWoSCCでの表記は、それぞれ、Inst Mol Sci、Inst Publ Hlthです。しかし、これらに含まれる語はいずれも機関名によく使われるため、類似した名称を持つ機関が多く、このまま検索すると大量のノイズを生ずるので注意が必要です。