

大学・公的機関名英語表記ゆれテーブル(Ver.2019.1)利用の手引き

2019年7月

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。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、データベースでは、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・分析しており、その成果をこのサイトから公開しています。それらのうち、「NISTEP 大学・公的機関名辞書」(以下「機関名辞書」と呼びます)と「大学・公的機関名英語表記ゆれテーブル」(以下「表記ゆれテーブル」、混同の恐れがなければ単に「テーブル」と略記します)のデータを更新し(機関名辞書は Ver.2018.2 → Ver.2019.1、表記ゆれテーブルは Ver.2018.1 → Ver.2019.1)、2019年7月に公開しました。

表記ゆれテーブルは、機関名辞書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 及び Scopus 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たものです(公開に当たり、WoSCC の提供元である Clarivate Analytics、及び Scopus の提供元である Elsevier 社の了解を得ております)。Ver.2019.1 では、機関名辞書由来のデータを更新するとともに、WoSCC と Scopus から抽出のデータも、2018年4月(WoS)と6月(Scopus)に NISTEP で行った機関同定に基づいて更新しています。Ver.2018.1 は、機関名辞書に含まれる名称と WoS に現れる表記を合体したものです。Scopus については、2013年に作成したデータを「大学・公的機関名英語表記ゆれテーブル(Scopus 版)(ver.2013.1)」として公開していましたが、Ver.2019.1 ではこの3つのデータ源(機関名辞書、WoS に現れる表記、Scopus に現れる表記)からのデータを合体しました。機関名辞書(Ver.2019.1)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¹。

¹ 機関名辞書 Ver.2019.1 については、このサイトの「NISTEP 大学・公的機関名辞書(ver.2019.1)利用マニュアル」をご覧ください。この公開版辞書では、機関の英語名は正式名称のみを収録し、別名はこのテーブルに移しております(正式名称は辞書とテーブルの両方に収録)。

※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.2019.1)の利用については、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.2019.1)

DOI: http://doi.org/10.15108/data_rsorg002_2019_1

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、国立研究開発法人等(特殊法人・独立行政法人を含む。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、代表機関だけでなく、主要な下部組織も含まれています。また、かつて存在したが統合、廃止、あるいは名称変更された非現存機関の主なものも収録しています。収録機関の詳細については、「NISTEP 大学・公的機関名辞書(ver.2019.1)利用マニュアル」の説明をご覧ください。

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、機関名辞書に含まれる機関表記(データ源が「辞書」となっているデータ)と WoSCC 及び Scopus から抽出した機関表記(データ源がそれぞれ「WoS」、「Scopus」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

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」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 サイト等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、NISTEP が整備を行っている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揺らぎ名データを収録しています。しかし、揺らぎ名データ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、Scopus 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。

このテーブルに含まれる機関名辞書由来の機関表記データは全部で 20,214、そのうち代表機関が 15,445、下部組織が 4,769 です。

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（National University Corporation, Independent Administrative Agency など）は原則として付いていません。大学下部組織の正式名称には、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 及び Scopus から抽出した機関英語表記データ

3.2.1 データ源

WoSCC については 1998-2017 年の期間、Scopus については 1996-2016 年の期間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を対象としました。これに該当する論文は、WoSCC では約 207 万件、Scopus では約 235 万件、その中の日本機関のデータは、WoSCC では延べ約 410 万件、Scopus では延べ約 443 万件になります。

3.2.2 テーブルに収録する機関表記の選択

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の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基準により選択しました（この基準は WoS と Scopus に共通）。

(1) 代表機関に同定された機関表記データのうち、次の(a)、(b)のどちらかの条件を満たすもの

(a) 出現頻度が 30 回以上

(b) 出現頻度が 5 回以上かつ同定された機関の中でその表記データの占有率が 20%以上

(2) 下部組織に同定された機関表記データのうち、次の(a)、(b)のどちらかの条件を満たすもの

(a) 出現頻度が 50 回以上

(b) 出現頻度が 5 回以上かつ同定された機関の中でその表記データの占有率が 20%以上

(1)、(2)いずれの場合も、出現頻度のみで選択を行うと大部分が大きな大学（あるいはその下部組織）のみになってしまうので、(b)の要素を加えています。

3.2.3 選択された機関表記の数

上記の方法により選択された表記バリエーションの数と、元々の出現頻度に対する抽出表記の出現頻度の比は次のようになります。

	同定された機関種	表記バリエーション数	頻度による抽出比
WoSCC	代表機関	4,429	96.8%
	下部組織	2,684	63.4%
Scopus	代表機関	4,685	64.6%
	下部組織	3,279	66.6%

従って、このテーブルにある機関表記を用いて WoSCC や Scopus の検索を行えば、対象の機関の 2/3 程度の論文データをカバーすることができます²。(テーブルにない表記でも、テーブル中の表記でカバーされるものもあるので、実際にはより高いカバー率になります。)

3.2.4 データベースの検索等にこのテーブルを利用される場合の留意事項

WoSCC や Scopus を用いて機関に関する検索やデータ分析を行う場合、このテーブルは有用ですが、留意していただくことがあります。これについては付録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 17,931(代表機関 14,348、下部組織 3,583)です。また、表記バリエーションの総数は 31,543(代表機関 21,332、下部組織 10,211)です。表記バリエーションの総数は、3.1 に示した機関名辞書由来の表記バリエーション数(20,214)、3.2 に示した WoSCC と Scopus の表記バリエーション数(それぞれ 7,113、7,964)の合計 35,291 より少ないですが、これらの二者または三者に共通の表記があるためです。

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. を参照)。

[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、Scopus のいずれか
- ・ WoSCC、Scopus における出現頻度 (データ源が WoSCC または Scopus 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

² 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。

- ・機関が属するセクター
- ・代表機関か下部組織かの別、及び代表機関の名称
- ・現存機関か非現存機関かの別（非現存機関の場合その継承機関があればその名称）

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、WoSCC、Scopus 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, Scopus:機関名辞書、WoSCC、Scopus のすべてに含まれる表記バリエーション
 - ・ 辞書, WoS:機関名辞書と WoSCC に含まれる表記バリエーション
 - ・ 辞書, Scopus:機関名辞書と Scopus 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
 - ・ Scopus:Scopus のみに含まれる表記バリエーション
- (5) WoS 頻度:データ源に「WoS」が含まれる場合、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。
- (6) Scopus 頻度:データ源に「Scopus」が含まれる場合、その表記バリエーションが Scopus 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、「NISTEP 大学・公的機関名辞書 (Version 2019.1)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。18 桁の固定長文字列で、先頭 3 文字は"NID"、後の 15 文字は数字です。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。

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。

	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
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。テーブルに含まれる非現存機関数は 5,913、表記バリエーション数は 9,955 です。
- (7) 継承機関名：「現状」が“**No**”の場合、その継承機関があればその名称を記入しています。
- (8) 補助セット：機関名辞書の一般的な機関収録基準には当てはまりませんが、データベース中の機関同定に必要なため例外的に収録する機関(主に下部組織)があります。そのような機関の場合、ここに“※”が付いています。詳しくは「**NISTEP 大学・公的機関名辞書 (Version 2019.1)利用マニュアル**」の 1.1(2)をご参照下さい。

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が、このテーブルをダウンロードされた場合、利用目的に従って自由に配列し直して下さい。

- 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(こ

の順に「機関内番号」を付番)

【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 を公開(英語別名等も含む)
- 2013/12/10 大学・公的機関名英語表記ゆれテーブル(WoSCC 版)(Ver.2013.1)、大学・公的機関名英語表記ゆれテーブル(Scopus 版)(Ver.2013.1) を公開
- 2014/11/14 NISTEP 大学・公的機関名辞書 Ver.2014.1 を公開(英語別名等も含む)
- 2014/11/14 大学・公的機関名英語表記ゆれテーブル(WoSCC 版)(Ver.2014.1) を公開
- 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(WoSCC 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 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)
- 2018/03/01 大学・公的機関名英語表記ゆれテーブル(Ver.2018.1)を公開(NISTEP 大学・公的機関名辞書の更新(Ver.2016.1→Ver.2018.1)及び2017年6月実施の WoSCC 機関同定の結果を反映)
- 2019/07/30 大学・公的機関名英語表記ゆれテーブル(Ver.2018.1)と大学・公的機関名英語表記ゆれテーブル(Scopus 版)(Ver.2013.1)を統合し、大学・公的機関名英語表記ゆれテーブル(Ver.2019.1)を公開(NISTEP 大学・公的機関名辞書の更新(Ver.2018.2→Ver.2019.1)及び2018年実施の WoSCC と Scopus の機関同定の結果を反映)

【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、下部組織名のみの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** 及び **Scopus** 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 及び Scopus の検索に利用する場合の留意事項

WoSCC や Scopus 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、その際に留意すべき点について述べます。

1. 著者所属機関のデータフィールド構成

提供機関から購入する XML データファイルでは、WoSCC、Scopus とともに、著者所属機関データはいくつかの XML 要素等(以降では、サブフィールドと記述します)に分割されています。

1.1 WoSCC のフィールド構成

次のサブフィールドがあります。

- (1) organization pref="Y": 標準化した代表機関名称 (多くの場合英語の正式名)
- (2) organization: 原則として代表機関の名称
- (3) suborganizattion: 原則として下部組織の名称
- (4) zip: 郵便番号
- (5) full_address: 詳しい機関・所在地情報
- (6) その他(city, state, country など)

organization の表記にはある程度のゆれがありますが、organization pref="Y" (オンラインデータベースでは「所属機関・拡張」)はそれぞれの代表機関に対してほぼ統一されていますので、これを用いて、高い再現率、精度で機関の検索や抽出をすることができます。organization pref="Y"による抽出結果と NITEP の名寄せ結果は、ほぼ一致することを確認しています。

上記の原則に従わず、suborganizattion に代表機関名、organization に下部組織名が記入されていることがあります。full_address、organization、country にはほぼすべてのレコードでデータが記入されていますが、他のサブフィールドは空白の場合があります。

1.2 Scopus のフィールド構成

次のサブフィールドがあります。

- (1) afid: 機関の識別番号
- (2) affiliation_org1: 機関名の第 1 階層
- (3) affiliation_org2: 機関名の第 2 階層
- (4) affiliation_org3: 機関名の第 3 階層
- (5) address-part: 所在地、郵便番号
- (6) その他(city, state など)

Scopus の特徴は、すべての機関データに(1)の機関識別番号が付けられていることです。非常に正確に付けられていますので、これを用いて、高い再現率、精度で機関の検索や抽出をすることができます。過去の検証作業で、Scopus の afid と NITEP の名寄せ結果は、ほぼ一致することを確認しています。

⁴ WoSCC や Scopus の書誌データの利用については、利用機関とそれぞれの提供機関 (WoSCC は Clarivate Analytics、Scopus は Elsevier 社) が取り交わす利用規約が適用されます。

Scopus では、organization という XML 要素の中に、所属機関の情報が含まれています。この organization 要素は、多くの場合、一つの住所情報に複数付与されています。ただし、各要素に含まれている組織情報の上下関係は書かれていませんので、ここでは(2)affiliation_org1、(3)affiliation_org2、(4)affiliation_org3という数字で表現しています。NISTEPの名寄せでは、上位3つまでの所属情報を抽出していますが、4つ以上の organization 要素を含む組織も存在しています。affiliation_org1～affiliation_org3 に含まれる情報はやや複雑で、原則として次のようになっています(この原則に従っていない場合もあります)。

- (i) すべてのサブフィールドにデータがある場合： affiliation_org1 に代表機関名、 affiliation_org2 にその次の下部組織名、 affiliation_org3 にその下の下部組織名
- (ii) affiliation_org2 と affiliation_org3 にデータがある場合： affiliation_org2 に代表機関名、 affiliation_org3 に下部組織名 (affiliation_org1 は空)
- (iii) affiliation_org3 のみにデータがある場合： affiliation_org3 に代表機関名(と下部組織名) (affiliation_org1 と affiliation_org2 は空)

従って、所属機関データを加工・分析するときは、次のようにするのが便利です。

- ・ (i)の場合はそのまま
- ・ (ii)の場合は affiliation_org2 のデータを affiliation_org1 に、 affiliation_org3 のデータを affiliation_org2 に移動。
- ・ (iii)の場合は affiliation_org3 のデータを affiliation_org1 に移動。

こうすれば、多くの場合、affiliation_org1 が代表機関名、affiliation_org2 が下部組織名になります。

このテーブルには、WoSCC の organization pref="Y"、Scopus の afid は含めていません。

なお、それぞれのデータベースのオンラインでの検索結果をダウンロードする場合は、WoSCC、Scopus のいずれも、著者所属機関は1つのフィールドに納められています。1論文に複数の所属機関が存在する場合、それぞれがセミコロン(;)で区切られています。

2. 機関名中の語の短縮形表記について

2.1 WoSCC

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(但しそれから揺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

Graduate School → Grad Sch

Institute → Inst

National → Natl

Science, Scientific → Sci

Medicine, Medical, Medicinal → Med

and → &

短縮形のあとにピリオド(.)は付きません。会社を示す"Co Ltd"、"Inc"などにもピリオドやカ

ンマは付きません。

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。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

2.2 Scopus

Scopus では、原則として短縮形表記は用いず、そのままの機関名、組織名が表記されます。前置詞や冠詞、ハイフンやアポストロフィーも省略されません。

但し、“Univ.”、“Dept.”、“Inst.”等の短縮形での表記も少数ですが見られます。University を“U.”と略する例も時々あります。これらは、古い年のデータに比較的多いようです。短縮語のあとにピリオドが付くことが多いですが、そうではないものもあります。“Co., Ltd.”が大多数ですが、“Co. Ltd.” “Co., Ltd” “Co. Ltd”なども見られます。

3. 代表機関と下部組織の表記

著者所属機関データのデータフィールド構成は1に述べたとおりですが、代表機関のサブフィールド、下部組織のサブフィールドに、それ以外の情報が含まれることがよくあります。以下、その主なものについて説明します。これは、WoSCC、Scopus にほぼ共通です。以下の表記例は短縮形を用いない Scopus 形式によりますが、WoSCC の場合は 2.1 に示した短縮形表記になります。

(1) 代表機関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。

- 代表機関名と下部組織名が合体した表記

[例] Hiroshima University Hospital(広島大学病院)

RIKEN BioResource Center(国立研究開発法人理化学研究所バイオリソースセンター)

- 所在地の一部が混入した表記

[例] Kyoto University Katsura(京都大学の桂キャンパス)

JAERI Takasaki(特殊法人日本原子力研究所の高崎研究所)

(2) 下部組織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は **Graduate School of Engineering** 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になります。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。

大学の下部組織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しばしばです。たとえば、東京大学工学系研究科には次のような表記バリエーションが含まれています。

- 専攻科(Department)の名称: Department of Applied Physics; Department of Mechanical Engineering など
- 研究科附属機関の名称: Photon Science Center(光量子科学研究センター); Quantum Phase Electronics Center(量子相エレクトロニクス研究センター)など

4. 表記揺れテーブルにおける WoSCC 及び Scopus 由来の表記の記述について

これまでに述べたように、WoSCC でも Scopus でも、代表機関と下部組織が記述されるサブフィールドや、その中の表記の方法は多様ですが、このテーブルでは、それらの表記を次のように整理しています。

(1) 原則から外れたサブフィールド記述の修正

各サブフィールドへの代表機関名、下部組織名の割り当てが1で述べた原則から外れている場合、できるだけ原則に従って置き換えて、代表機関表記、下部組織表記を抽出します。

(2) 代表機関サブフィールド中の下部組織混合表記

代表機関サブフィールドに、3(1)に例示したような代表機関と下部組織が合体した表記(“Hiroshima University Hospital”など)が記入されている場合、下部組織の表記とみる方が適切なのでそちらに移します。

(3) 代表機関、下部組織以外の情報が混入した表記

所在地の一部が混入した表記や、機関名辞書の収録対象下部組織(原則として代表機関直下の組織)以外の下部組織名が混入した表記では、それらを除去します。たとえば、3(1)に例示した“Kyoto University Katsura”は“Kyoto University”に、“Graduate School of Science, Department of Chemistry”は“Graduate School of Science”に修正します。

(4) 大学直下の下部組織名が省略されてその下の組織が記入されている場合

3(2)で述べたように、大学の下部組織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よくあります。NISTEPの機関同定プログラムでは、大学の主要な低層下部組織表記を、機関名辞書に収録されているその上位の下部組織に同定する仕組みがあります。このテーブルにはその種の表記も含まれます。

たとえば、代表機関サブフィールドの表記が“University of Tokyo”で下部組織サブフィールドの表記が3(2)に例示した“Department of Applied Physics”である場合、東京大学では Department of Applied Physics は工学系研究科(Graduate School of Engineering)に属する可能性が極めて高いことが他のデータから知られているので、この表記をそこに含めます。

(5) Scopus におけるピリオドの除去

2.2 で述べたように、Scopus では短縮形の後のあとにピリオド(.)が付く場合と付かない場合があります。このテーブルでは、ピリオドをすべて外して両方の表記を統合しました。また、”Co., Ltd.”等のピリオドとカンマをすべて除去し、”Co Ltd”に統一しました。その他の記号(ハイフン、アポストロフィー等)はそのままにしています。

(6) 出現頻度の算出

(1)~(5)の処理を行う過程で、代表機関、下部組織ともに同一の表記が合体されていきます。テーブルの「WoS 頻度」、「Scopus 頻度」は、合体のたびに頻度を合計したものです。

5. 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(その後国立研究開発法人)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 と表記されることとなりますが、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(Tokyo University of Agriculture and Technology)と東京農業大学(Tokyo University of Agriculture)、浜松医科大学(Hamamatsu University School of Medicine)と浜松大学(Hamamatsu University)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)の正式名称は、それぞれ、Institute for Molecular Science、The Institute of Public Health 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。