

Scopus-NISTEP 大学・公的機関名辞書対応テーブル 説明書

2013 年 7 月 1 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、たとえば東京大学の正式英語名は(The) University of Tokyo ですが、これが University Tokyo、Tokyo University などと表記されたり、“University”が”Univ.”、“Univ”、“U.”などと略記されたりします。

この問題は、データベースに含まれる機関名データがどの機関を表しているかを正しく同定できれば解決されます。今回、科学技術・学術政策研究所(NISTEP)では、世界最大級の書誌・引用データベースである Scopus(Elsevier 社製)に含まれる機関名データから、国内の大学・公的機関(どのような機関を含むかについては2. (2)をご覧ください)の機関同定を行いました。具体的には、Scopus の機関名データを、NISTEP 大学・公的機関名辞書(以下、「機関名辞書」)の収録機関に対応させる「Scopus-NISTEP 大学・公的機関名辞書対応テーブル」を作成しました。Scopus データベースの利用や、国内機関の論文生産に関する調査分析に役立てていただくことを念頭に、エルゼビア・ジャパン株式会社の了解を得て、このテーブルを公開いたします。

なお、このサイトから既に公開している以下のデータも併せてご利用下さい。

- NISTEP 大学・公的機関名辞書データ:10,000 以上の国内機関の和英の名称、属するセクター、変遷情報(統廃合、改称等)等を収録した辞書データです。大学、公的機関が中心ですが、研究活動を行っているそれ以外の機関もできるだけ収録しています。(説明資料:NISTEP 大学・公的機関名辞書利用マニュアル)
- 大学・公的機関名英語表記ゆれテーブル:研究論文が多い約 200 の国内大学と約 50 の公的機関について、Scopus によく出現する機関名表記のゆれをまとめたデータです。(説明資料:大学・公的機関名英語表記ゆれテーブル利用の手引き)

※このテーブルの利用について

Scopus-NISTEP 大学・公的機関名辞書対応テーブルの利用については、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-非営利」を適用します。すなわち、以下の条件に従う場合に限り、1)本テーブルを複製、頒布、展示、実演、2)二次的著作物を作成することができます。



表示 - あなたは原著作者のクレジットを表示しなければなりません。

非営利 - あなたはこの作品を営利目的で利用してはなりません。

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: Scopus-NISTEP 大学・公的機関名辞書対応テーブル

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. 同定の対象と方法

(1) 同定対象のデータ

今回同定を行ったデータは、1996-2010年の期間に Scopus データベースに採録された論文の著者所属機関データのうち、日本の機関と判別されたデータです。この期間に、日本の機関を含む論文は約 150 万件、その中の日本機関のデータは延べ 271 万件存在します。

(2) 機関の属するセクターによる同定のレベル

機関名辞書では、機関を 16 のセクターに分類しています。この同定プロジェクトは、NISTEP の「大学・公的機関における研究開発に関するデータ整備」の一環として行っているため、大学・公的機関に属するセクターの機関の同定に主眼を置いています。16 のセクターを大学・公的機関とそれ以外の機関に分けると次の通りです。

| 区分 | 属するセクター |
|---------|--|
| 大学・公的機関 | 国立大学;国立短大;国立高専;公立大学;公立短大;公立高専;大学共同利用機関;国の機関;特殊法人・独立行政法人;私立大学;私立短大;私立高専 |
| それ以外の機関 | 地方公共団体の機関;会社;非営利団体;その他の機関 |

「大学・公的機関」については機関レベルでの同定を行っていますが、「それ以外の機関」については属するセクターのみを同定し、機関同定は行っていません*1。このテーブルの全レコードのうち約 80%は「大学・公的機関」に属しています。

(3) 同定レベル

Scopus の各機関名データに対し、同定の精確度を、次の 5 つのレベルで示しています。

| 同定レベル | 説明 |
|-------|-------------------|
| I | かなり高い確度で機関を同定。 |
| U | やや低い確度で機関を同定。 |
| S | 機関を同定せず、セクターのみ同定。 |
| N | 国内機関であることのみ同定。 |

(2)で述べた考え方により、「大学・公的機関」の同定レベルは”I”または”U”ですが(”I”が 99%以上)、まれに”S”のものがあります。また、「大学・公的機関」に同定されていない機関レコード(全体の約 20%)のうち約 2/3 の同定レベルは”S”(2)の「それ以外の機関」に属するセクターのいずれかに分類)、残りは同定レベル”N”(国内機関であることだけが判明)です(末尾の【参考】のデータを参照)。

*1 内部的には大部分の「それ以外の機関」に対しても機関同定を行っていますが、精確度にやや問題があるため、公開はしておりません。

3. テーブルの各項目

Scopus-NISTEP 大学・公的機関名辞書対応テーブルには、約 270 万の著者所属機関レコードが存在します。テーブルの各項目について説明します。

- (a) Scopus 記事番号(scopus_eid): 当該機関を著者所属機関に含む Scopus の記事番号です。
- (b) Scopus 記事内の著者所属機関番号(scopus_address_seq): 1 つの Scopus 記事 (scopus_eid が同一) の中に存在する著者所属機関レコードの中での当該レコードの順番です。最初のレコード番号が 0、以下 1, 2, …となります。日本以外の所属機関のレコードはこのテーブルに含まれていませんので、1 つの scopus_eid に対しすべての順番が存在するとは限りません。
- (c) 同定レベル(match_level): 2(3)で述べた同定の精確度を示す記号で、I, U, S, N のいずれかです。
- (d) 同定機関の機関 ID(nid): 同定された機関に機関名辞書で与えられている識別番号です。この番号を用いて、機関名辞書により機関の和・英の名称、所属セクター、その他の情報を得ることができます。
- (e) 同定機関セクター(sector): 同定された機関が属するセクターで、2(2)に示す 16 のいずれかです。

レコードの同定レベルによって、上記(a)～(f)のデータ存否は次のようになります。

| テーブルの項目 | | | | | |
|---------|-----|-----|-----|-----|-----|
| 同定レベル | (a) | (b) | (c) | (d) | (e) |
| I | ○ | ○ | ○ | ○ | ○ |
| U | ○ | ○ | ○ | ○ | ○ |
| S | ○ | ○ | ○ | | ○ |
| N | ○ | ○ | ○ | | |

4. このテーブルの利用法

このテーブルは、主に次の 2 つの利用法が考えられます。

(1) Scopus での著者所属機関検索・分析の補助ツールとして

これには次の二通りの利用が考えられます。

第一は、Scopus で検索した論文データ集合における所属機関(大学または公的機関)の同定(名寄せ)です。Scopus のカスタムデータを用いる場合は、検索したデータに scopus_eid の項目があります。これらの scopus_eid を「Scopus-NISTEP 大学・公的機関名辞書対応テーブル」の scopus_eid と接合することで、機関名の名寄せが可能となります。Scopus のオンラインデータを用いる場合は、「7. scopus_eid の取得方法について」に示した方法で、分析対象とする論文データについての scopus_eid を抽出してください。抽出された scopus_eid を、同様にこのテーブルと接合します。

第二の利用方法は、ある機関の論文データの一括検索です。まず、検索したい機関の機関 ID を機関名辞書で調べます。次に、このテーブルを用いてその機関 ID を持つ論文データに対する scopus_eid の集合を作り、Scopus データベースからそれらに一致するレコードを抽出します。

これにより、Scopus 中の機関名表記のゆれに関わりなく、漏れのない機関検索が行えます。但し、この方法は Scopus のカスタムデータにのみ適用できます。オンラインデータをご使用の方は、このサイトで公開している「大学・公的機関名英語表記ゆれテーブル」によって検索したい機関の表記バリエーションが得られるので、それらを用いて機関名の OR 検索を行って下さい。

(2) 国内機関の論文生産統計の基礎データとして

このテーブルと機関名辞書を用いて、1996-2010 年の期間における大学・公的機関の論文生産統計をとることができます。また、大学・公的機関以外の機関も含めたセクター別の論文生産統計も得ることができます。

但し、レコードを単純に集計した結果は、機関またはセクターの合計論文数ではなく、Scopus データベースに出現した著者所属機関レコードの合計数であることにご注意下さい。一つの論文に同じ機関の異なる部局の著者が含まれている場合、この機関のレコードが複数存在する（それぞれ部局が異なる）ことがあります。論文数の統計をとる場合には、同じ scopus_eid の中の同じ機関（機関 ID が同じ）のレコードの重複を削除する必要があります。

scopus_eid を用いると、異なる機関あるいは異なるセクターの間でどれくらい共著論文があるか（共同研究が行われているか）を調べることもできます。

なお、このテーブルで可能なのは、1996-2010 の全期間にわたる統計だけです。期間、分野、ドキュメントの種類を区切った統計を得るには、Scopus データベースと情報を組み合わせる必要があります。

5. 利用上の注意

(1) 下位部局の上位機関への統合

機関名辞書には、いくつかの機関の下位部局も登録され、それらには上位機関とは別の機関 ID が与えられています。しかし、このテーブルでは、同定対象を上位機関に統一しています。機関名辞書で「代表機関フラグ」が FALSE である機関は下位部局ですので、その上位機関の機関 ID（機関名辞書で容易に判ります）によりこのテーブルを参照して下さい。

但し、大学共同利用機関である自然科学研究機構、情報・システム研究機構、人間文化研究機構の 3 機関に限り、その下部機関の研究所（分子科学研究所、統計数理研究所、国立民族学博物館等）単位で同定を行っています（一部上位の機構単位で同定されているデータがあります）。

(2) Scopus の機関名レコードに 2 つの機関が含まれている場合の措置

次の例のように、Scopus の 1 つの機関名レコードに 2 つの機関が含まれていることがあります。

| |
|--|
| Department of Molecular and Developmental Biology National Institute of Genetics The Graduate University for Advanced Studies (SOKENDAI) |
|--|

この例では、国立遺伝学研究所と総合研究大学院大学という 2 つの異なる機関が 1 つの機関名レコードに記載されています。このような例は、主に一人の著者が異なる機関に属する場合

に見られるようです。

このような場合、このテーブルでは、2つの機関に対応するデータをセミコロン(;)で区切って記入しています。上記の例では次のようになります。

| |
|--|
| 同定レベル:I;I |
| 同定機関の ID:NID201200921752444;NID201200135086250 (これらはそれぞれ国立遺伝学研究所と総合研究大学院大学の機関 ID です。) |
| 同定機関のセクター:大学共同利用機関;国立大学 |

但し、このような識別ができず、どちらか一方の機関のみ同定されているデータもあります。

6. 注記

① Scopus-NISTEP 大学・公的機関名辞書対応テーブルの精度

このテーブルの作成には十分な注意を払っておりますが、すべての同定結果を人手でチェックはしていませんので、少数の同定エラーがあります。サンプルデータのチェックの結果では、同定レベル”I”のレコードのエラー率は 0.3%未満ですが、同定レベル”U”のレコードでは 10%強、全体では 0.4%未満と推定されます。

これらについては、同定アルゴリズムの精密化、機関名辞書のデータ充実等により改善を行っていく予定です。ご使用に当たって注意下さるとともに、お気づきの点をお知らせ下さると幸いです。

| セクター | 同定レベル | | | | Total |
|-------------|-----------|--------|---------|---------|-----------|
| | I | U | S | N | |
| 国立大学 | 1,249,340 | 5,297 | | | 1,254,637 |
| 国立短大 | 440 | 15 | | | 455 |
| 国立高専 | 10,821 | 222 | | | 11,043 |
| 公立大学 | 122,507 | 2,253 | | | 124,760 |
| 公立短大 | 919 | 54 | | | 973 |
| 公立高専 | 935 | 23 | | | 958 |
| 大学共同利用機関 | 32,710 | 313 | | | 33,023 |
| 国の機関 | 49,367 | 1,131 | 360 | | 50,858 |
| 特殊法人・独立行政法人 | 236,073 | 4,103 | | | 240,176 |
| 地方公共団体の機関 | | | 53,969 | | 53,969 |
| 私立大学 | 437,789 | 6,493 | 11 | | 444,293 |
| 私立短大 | 2,387 | 350 | | | 2,737 |
| 私立高専 | 148 | 4 | | | 152 |
| 会社 | | | 259,262 | | 259,262 |
| 非営利団体 | | | 58,619 | | 58,619 |
| その他 | | | 691 | | 691 |
| セクター不明 | | | | 183,303 | 183,303 |
| 合計 | 2,143,436 | 20,258 | 372,912 | 183,303 | 2,719,909 |

【参考】セクター別、同定レベル別の機関レコード数*2

*2 5(2)で述べた2機関を含むレコードについては、2つに分割して集計しています。

② Scopus-NISTEP 大学・公的機関名辞書対応テーブルのカバー率

Scopus-NISTEP 大学・公的機関名辞書対応テーブルのもととなるデータは、2010 年末時点の Scopus カスタムデータです。したがって、出版年が 2011 年以降の論文については、本テーブルには含まれていません。また、Scopus では、適時データの追加、修正が行われていることから、2010 年以前についても、カバー率は 100%とはなっていません。

以下に 2013 年 7 月 1 日時点の Scopus に含まれている日本論文数(全分野の全てのドキュメントタイプを対象)と Scopus-NISTEP 大学・公的機関名辞書対応テーブルがカバーする論文数の比較を示します。

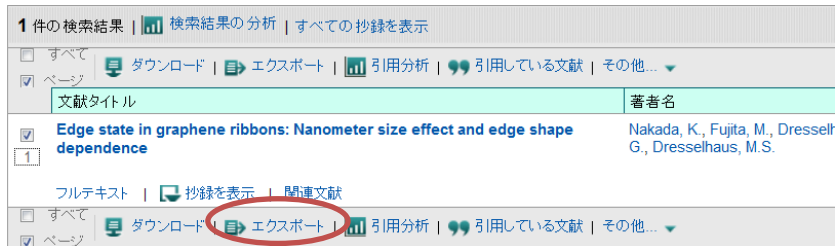
2009 年までは 90%を超えています。2010 年のカバー率は約 85%、2011 年以降は 0%です。このように Scopus-NISTEP 大学・公的機関名辞書対応テーブルのカバー率は年毎に変化しますので、本テーブルを利用して分析を行う際は、カバー率を必ず調べるようにしてください。

| 出版年 | Scopus中の日本論文 (2013年7月1日時点) | Scopus-NISTEP大学・公 的機関名辞書対応テー ブル(2010年末時点) | カバー率 |
|------|-------------------------------|---|-------|
| 1996 | 84,493 | 82,373 | 97.5% |
| 1997 | 90,569 | 85,909 | 94.9% |
| 1998 | 89,756 | 86,738 | 96.6% |
| 1999 | 92,555 | 89,484 | 96.7% |
| 2000 | 94,171 | 90,647 | 96.3% |
| 2001 | 91,933 | 88,601 | 96.4% |
| 2002 | 92,637 | 88,660 | 95.7% |
| 2003 | 99,676 | 96,083 | 96.4% |
| 2004 | 107,013 | 96,707 | 90.4% |
| 2005 | 116,185 | 105,507 | 90.8% |
| 2006 | 118,161 | 108,177 | 91.6% |
| 2007 | 113,894 | 105,010 | 92.2% |
| 2008 | 113,996 | 105,532 | 92.6% |
| 2009 | 118,551 | 110,979 | 93.6% |
| 2010 | 119,778 | 101,573 | 84.8% |
| 2011 | 121,356 | 36 | 0.0% |
| 2012 | 119,091 | 0 | 0.0% |
| 2013 | 50,229 | 0 | 0.0% |

【参考】Scopus-NISTEP 大学・公的機関名辞書対応テーブルのカバー率

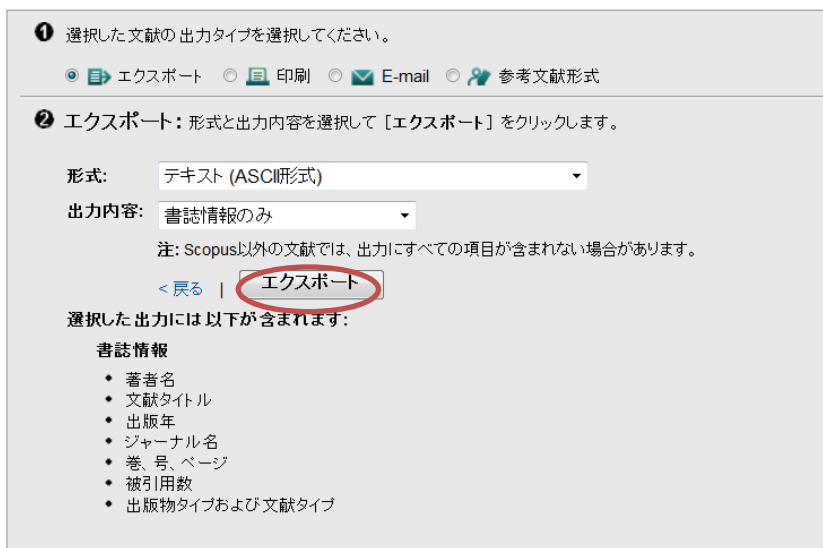
7. scopus_eid の取得方法について

- scopus_eid は、Scopus に収録されている論文についてのユニークな ID 番号です。Scopus の検索画面(<http://www.scopus.com/home.url>)から、次の方法で取得可能です。なお、Scopus を利用するには、エルゼビア・ジャパン社との契約が必要です。
- Scopus の検索で得られた論文の内、書誌情報等をダウンロードしたい論文にチェックをつけ、「エクスポート」を選択してください。



- 次にエクスポートの形式と出力内容を選んでください。この時、書誌情報を出力内容に必ず含めてください。「エクスポート」ボタンを押下するとデータが出力されます。

出力: エクスポート、印刷、E-mail、参考文献形式



- ダウンロードした書誌情報には、Scopus 上の該当論文の URL の情報が含まれています。この URL の中で、下線を引いた部分が scopus_eid に該当します。形式として、コンマ区切りファイルを選ぶことで、エクセルで編集可能なデータがダウンロードできます。

Scopus
エクスポート日付:2013/04/24

Nakada, K., Fujita, M., Dresselhaus, G., Dresselhaus, M.S.
Edge state in graphene ribbons: Nanometer size effect and edge shape dependence
(1996) Physical Review B - Condensed Matter and Materials Physics, 54 (24), pp. 17954-17961. 被引用数 1496 回.
<http://www.scopus.com/inward/record.url?eid=2-s2.0-0000781318&partnerID=40&md5=2fca1ab640dfce62a4196d006582e9d4>
文献タイプ: Article
情報源: Scopus