

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—

2015 年 10 月

文部科学省 科学技術・学術政策研究所

科学技術・学術基盤調査研究室

小野寺 夏生 伊神 正貫 阪 彩香

NISTEP NOTE(政策のための科学)は、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」に関する調査研究やデータ・情報基盤の構築等の過程で得られた結果やデータについて、速報として関係者に広く情報提供するために取りまとめた資料です。

NISTEP NOTE (Science of Science, Technology and Innovation Policy) No. 15

Preparation and Utilization of
the NISTEP Dictionary of Names of Universities and Public Organizations
– Toward Research Data Analysis of Universities at a Subsidiary Organization Level –

October 2015

Research Unit for Science and Technology Analysis and Indicators,
National Institute of Science and Technology Policy (NISTEP)
Ministry of Education, Culture, Sports, Science and Technology (MEXT)
Japan

本報告書の引用を行う際には、出典を明記願います。

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―

小野寺 夏生, 伊神 正貫, 阪 彩香¹

¹文部科学省 科学技術・学術政策研究所 科学技術・学術基盤調査研究室

要旨

科学技術・学術政策研究所(NISTEP)では、研究開発を実施している我が国の機関の基本的情報(和英の名称、所属セクター、主要な下部組織、機関の変遷等)を収録する「NISTEP大学・公的機関名辞書」の作成、維持、公開に取り組んでいる。本報告書では、機関名辞書の最新の整備状況について報告する。特に、最新の公開バージョンであるNISTEP大学・公的機関名辞書(Version 2015.1)において拡充を行った、機関名辞書への大学下部組織情報の付加について詳しく説明する。また、大学下部組織情報を整備する過程で明らかになった、最近の大学組織構造の傾向についても述べる。加えて、機関名辞書の活用事例として、機関名辞書を用いて論文データベースの機関同定を行うためにNISTEPで開発を行っている機関同定プログラムの説明とそのプログラムを用いた機関同定結果についても紹介する。

Preparation and Utilization of the NISTEP Dictionary of Names of Universities and Public Organizations

- Toward Research Data Analysis of Universities at a Subsidiary Organization Level -

Natsuo Onodera, Masatsura Igami, and Ayaka Saka¹

¹ Research Unit for Science and Technology Analysis and Indicators, National Institute of Science and Technology Policy (NISTEP), MEXT

ABSTRACT

National Institute of Science and Technology Policy (NISTEP) has prepared, maintained, and made public the “NISTEP Dictionary of Names of Universities and Public Organizations,” which contains basic information of Japanese research organizations such as Japanese and English names, the sector to which organizations belong, principal subsidiary organizations, and defunct organizations. This report describes the present state of the Dictionary, particularly addressing the information on subsidiary organizations of universities enhanced in the recent version of the Dictionary (Version 2015.1). Based on the investigation into those subsidiaries, recent trends of complication and diversification of organizational structure in universities are described.

As one of applications of the Dictionary, NISTEP has performed institution name disambiguation for the author affiliation data in bibliographic databases. This note outlines the development of an institution-name-disambiguation system by NISTEP and the results of application of this system.

(裏白紙)

目次

概要

1 NISTEP 大学・公的機関名辞書の概要	1
1-1 コアセットに含まれる代表機関	2
1-2 コアセットに含まれる下部組織	2
1-3 コアセットに含まれる情報.....	3
2 最近の大学組織構造の傾向.....	4
2-1 学部、大学院の構成の多様化・複雑化	5
2-2 全学的組織の拡大.....	5
2-3 頻繁な組織変更.....	5
3 NISTEP 大学・公的機関名辞書を用いた論文データベースの機関同定	6
3-1 全体的な結果	6
3-2 32 大学の下部組織同定	7
4 まとめと今後の課題	10

本編

1 背景と目的.....	11
2 NISTEP 大学・公的機関名辞書の継続的整備.....	13
2-1 大学・公的機関に関する基本的情報のアーカイブ化.....	13
2-2 機関名辞書への下部組織レベル情報の付加の必要性	13
2-3 機関名辞書の概要.....	14
2-4 機関名辞書に収録するデータの収集	17
3 NISTEP 大学・公的機関名辞書への大学下部組織情報の付加	20
3-1 下部組織情報を網羅的に収録する大学とその選定方法.....	20
3-2 下部組織情報の収集の実際	22
3-3 最近の大学組織構造の傾向	24
3-4 大学下部組織情報の収集の難しさ	30
4 NISTEP 大学・公的機関名辞書を用いた論文データベースの機関同定	32

4-1 論文書誌情報(著者所属機関表記)に基づく機関同定の概要とその難しさ	32
4-2 大学の下部組織同定の問題点.....	36
4-3 機関同定アルゴリズムの骨子.....	37
4-4 機関名辞書の補助セット.....	40
4-5 同定漏れや誤同定を防ぐための具体例.....	41
4-6 大学の下部組織同定率向上策.....	43
5 論文の機関同定の結果 – 大学下部組織同定を中心に –	46
5-1 全体的な結果	46
5-2 32 大学の下部組織同定	48
5-3 目視による同定結果の検証.....	53
5-4 機関名の同定精度の検証.....	55
6 まとめと今後の課題	56
参考文献.....	58
調査担当	59

概要

(裏白紙)

概要

科学技術・学術政策研究所(NISTEP)では、研究開発を実施している我が国の機関(大学と公的機関に重点)の基本的情報を収録する「NISTEP 大学・公的機関名辞書」の作成、維持、公開に取り組んでいる。本報告書では、機関名辞書の最新の整備状況について報告する。特に、最新の公開バージョンである NISTEP 大学・公的機関名辞書(Version 2015.1)において拡充を行った、機関名辞書への大学下部組織情報の付加について詳しく説明する。また、大学下部組織情報を整備する過程で明らかになった、最近の大学組織構造の傾向についても述べる。加えて、機関名辞書の活用事例として、機関名辞書を用いて論文データベースの機関同定を行うために NISTEP で開発を行っている機関同定プログラムの説明とそのプログラムを用いた機関同定結果についても紹介する。本報告書のポイントは以下の通りである。

1 NISTEP 大学・公的機関名辞書の概要

NISTEP 大学・公的機関名辞書(以下「機関名辞書」という)には、研究開発を行っている日本の機関(以下「代表機関」という)とその主な下部組織の名称(日本語と英語)、所属するセクター、階層関係、変遷等の情報を収録している。これらの情報は、コアセットと補助セットの 2 つのデータセットに分けて収録している。

概要図表 1 にコアセットと補助セットの関係を示す。コアセットには機関名辞書の基本情報が収録されており、NISTEP の「データ・情報基盤」の一つとして公開している¹。補助セットには、コアセットへの収録基準に合致しないが、論文データベース中の所属機関データの同定や分析の作業上収録の必要性が高い機関や情報を収録している。これらは、NISTEP 内部で分析を行う際に用いるものであるが、一部については、「大学・公的機関名英語表記ゆれテーブル」として公開している。

概要図表 1 機関名辞書のコアセットと補助セット

コアセット

- 網羅的なリストの継続的な収集・アーカイブ化
- データ・情報基盤として公表

補助セット

- 論文データベース(Web of Science, Scopus)等の機関名同定に必要な補助的な情報の蓄積
- NISTEP内の分析で利用(機関名辞書を用いて論文データベースの名寄せを行うプログラムも併せて開発)
- 分析結果や分析の過程で得られたデータ(大学・公的機関名英語表記ゆれテーブル等)を公表

¹ 最新の公開バージョンは、NISTEP 大学・公的機関名辞書(Version 2015.1)であり、本報告書と同時に公開を行っている。

1-1 コアセットに含まれる代表機関

機関名辞書(コアセット)に含まれる代表機関は、つぎの機関である。

機関名辞書(コアセット)に含まれる代表機関

- ① 現存する全ての大学、短期大学、高等専門学校、大学共同利用機関及び学校法人
- ② 研究開発を行っている国の機関、特殊法人・独立行政法人²等、地方公共団体の機関、非営利法人、その他の機関(日本所在の国際機関等)
- ③ ①または②のセクターに属し、この15年ほどの間に統廃合、名称変更等があつて現存しない機関のうち主要なもの

民間企業はコアセットでなく補助セットに収録している。その理由は、別途に構築、公開しているNISTEP 企業名辞書に、我が国の企業に関する主要な情報が収録されているためである。しかし、論文データベースには企業から発表される論文も多数含まれているので、補助セットを用いてそれらの同定、分析を可能としている。

1-2 コアセットに含まれる下部組織

研究開発の動向を把握するとき、下部組織レベルでの調査・分析が望ましい機関がある。大規模な総合大学では、いろいろな分野の研究が多数行われているので、詳細な研究活動の状況を把握するには、その下部組織(学部、大学院研究科、附置研究所等)ごとの情報が重要である。また、大学共同研究機関や一部の独立行政法人では、元々独立していた機関を「機構」という名の統合組織にまとめているが、これらについても、下部組織ごとの分析が意味を持つ場合が多い。このようなことから、機関名辞書では重要な下部組織も収録している。機関名辞書(コアセット)には、以下の下部組織を収録する。現存しない下部組織も含む。原則としては第2階層下部組織(大学の直下の下部組織)までを対象とする。

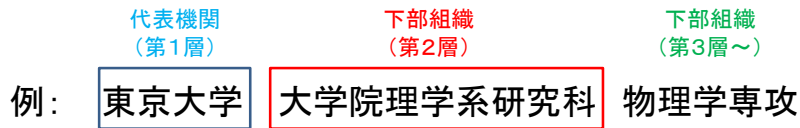
概要図表 2 に代表機関-第2階層下部組織-第3階層下部組織の関係性のイメージを示した。

機関名辞書(コアセット)に含まれる下部組織

- ① 研究論文数の特に多い大学の下部組織。当面、概要図表 3 に示す32大学を対象とし、それらの第2階層下部組織を網羅的に収録した(事務組織を除く)。病院及び共同利用・共同研究拠点となっている下部組織については第3階層下部組織(医学部の附属病院など)であっても収録対象とした。
- ② その他の大学に属する共同利用・共同研究拠点となっている下部組織。
- ③ 大学共同利用機関である4つの機構に直属する研究所等。
- ④ 国の機関及び特殊法人・独立行政法人等に属する病院、大学校、主要な研究所等。

² ここに挙げる独立行政法人の中には、2015年4月1日に国立研究開発法人に移行したものが含まれているが、本報告書の記述は機関名辞書の最新公開バージョン(Version 2015.1)に基づいており、それは2015年1~2月時点の調査によるものであるため、ここではすべて独立行政法人とする。

概要図表 2 代表機関-第2階層下部組織-第3階層下部組織



概要図表 3 機関名辞書に下部組織を網羅的に登録する 32 大学

種別	大学	下部組織数*	備考**	種別	大学	下部組織数*	備考**
国立大学	北海道大学	86	○	国立大学	大阪大学	64	○
	東北大学	79	○		神戸大学	42	○
	筑波大学	67	○		岡山大学	55	
	群馬大学	21			広島大学	53	○
	千葉大学	48	○		徳島大学	39	
	東京大学	74	○		九州大学	134	○
	東京医科歯科大学	26			長崎大学	35	
	東京工業大学	67	○		熊本大学	38	
	東京農工大学	24		公立大学	大阪市立大学	35	
	新潟大学	27			大阪府立大学	24	
	富山大学	36		私立大学	慶應義塾大学	53	
	金沢大学	50			東海大学	71	
	信州大学	30			東京理科大学	31	
	岐阜大学	27			日本大学	51	
	名古屋大学	67	○		早稲田大学	90	
	京都大学	70	○		近畿大学	56	

* 機関名辞書収録下部組織数 (Version 2015.1に収録の現存する組織)

** ○印は2011年度から、それ以外は2012年度から収録。

1-3 コアセットに含まれる情報

機関名辞書の最新版 (Version 2015.1) のコアセットに収録されているセクター別の機関数を概要図表 4 に示す。機関名辞書 (コアセット) には、1-1 及び 1-2 に該当する代表機関及び下部組織に対する、下に示した情報を収録する。また、現在公開している機関名辞書 (Version 2015.1) の情報の取得時期及び非現存機関のカバー範囲を概要図表 5 に示す。従って、この取得時期以降の機関、組織の新設や変遷の情報は含まれていない。

機関名辞書(コアセット)に含まれる情報

- ① NISTEP が与える機関 ID (代表機関、下部組織のそれぞれに独立して与える)
- ② 日本語と英語の正式名
- ③ 所属するセクター (17 に分類、概要図表 4 を参照)
- ④ 下部組織とそれが属する代表機関とのリンク
- ⑤ 現存しない代表機関、下部組織の変遷情報 (移行の種別、移行年月日、継承先がある場合その機関)
- ⑥ 郵便番号

概要図表 4 機関名辞書(コアセット)の収録機関数(Version 2015.1)

セクター	セクター	現存機関			非現存機関			合計		
		代表機関	下部組織	小計	代表機関	下部組織	小計	代表機関	下部組織	小計
1	国立大学	86	1,268	1,354	15	122	137	101	1,390	1,491
2	国立短大				26		26	26		26
3	国立高専	51		51	8		8	59		59
4	公立大学	85	60	145	16	1	17	101	61	162
5	公立短大	17		17	45		45	62		62
6	公立高専	3		3	4		4	7		7
7	大学共同利用機関	4	24	28	3		3	7	24	31
8	国の機関	40	9	49	57	7	64	97	16	113
9	特殊法人・独立行政法人	88	224	312	37	29	66	125	253	378
10	地方公共団体の機関	685		685	90		90	775	0	775
11	学校法人	663		663	1		1	664		664
12	私立大学	612	364	976	41	17	58	653	381	1,034
13	私立短大	341		341	223		223	564		564
14	私立高専	3		3				3		3
15	会社									
16	非営利団体	3,568		3,568	318		318	3,886		3,886
17	その他	5		5				5		5
	総計	6,251	1,949	8,200	884	176	1,060	7,135	2,125	9,260

概要図表 5 機関名辞書(Version 2015.1)の情報の取得時期及び非現存機関のカバー範囲

機関の種別	最新情報取得時期	非現存機関のカバー範囲*
大学等(短大、高専、大学共同利用機関を含む)	2015年1月	2008年度以降
32大学の下部組織	2015年2月	2013年度以降
国の機関、特殊法人・独立行政法人等	2015年1月	2008年度以降

* 確実にカバーしている範囲であり、それ以前に現存しなくなった機関も含まれる。

2 最近の大学組織構造の傾向

2000年以前の大学の教員組織は、学部と大学院研究科、それに附置研究所が中心であった。学部の下には学科、学部附置研究所、学部附属施設(病院等)があり、また大学院研究科の下には専攻があり、それらの名称の付け方は、大学間でかなり共通性があった。もちろん、現在でもこの構造は保持されている場合もあるが、かなり大きな変動があり、大学の組織構成が多様・複雑化している。このため、上述した「第2階層までの下部組織を網羅的に収録」の方針に従って下部組織を選別することは必ずしも容易でない。

しかし、このような困難性があるからこそ、どこかで一元化された情報がアーカイブ化されていないと、必要な調査や分析が行われなくなることになる。NISTEPの機関名辞書は、大学下部組織情報のアーカイブ化を一つの重要な目標としている。そのためには、変遷する下部組織の情報(どの学部・研究科等が、いつ、どの学部・研究科等に変ったか)を不断に蓄積する必要がある。以下に、最近の大学組織構造の傾向を示す。

2-1 学部、大学院の構成の多様化・複雑化

(1) 「学部」、「研究科」を表す呼称の多様化と教員組織の設置

4年制を表す「学部」の代わりに学類、学域等を用いる大学、また大学院を表す「研究科」の代わりに院、研究院、学府等の呼称を用いる大学が出現している。また、最近では、学部や研究科以外に教員組織を置き、教員はそこに属して学部や大学院の教育を担当するという形にする大学が出現している(北海道大学、筑波大学、金沢大学、信州大学等)。その呼称は「系」、「研究院」、「研究域」など様々である。

(2) 階層関係の複雑化

金沢大学は2008年に、8つの学部を3学域(人間社会、理工、医薬保健)、16学類に再編した。16の学類は従来の学部よりはやや細分されているが学部相当とみられる(学類の下に、学科にはほぼ相当する「コース」が置かれる)。早稲田大学では、分野別の10の「学術院」があり、各学術院の下に、学部、大学院研究科、研究所が存在する。たとえば理工学学術院には3つの学部と6つの研究科が属する。金沢大学の学域と早稲田大学の学術院は性格を異にするが、いずれも代表機関と第2層下部組織の間に位置する「第1.5階層」ともいべき組織である。

(3) 学部の名称の多様化

1998年度と2014年度について32大学の学部名称を比較すると、約40名称増加しており、この期間に学部の名称が大きく多様化したことが明らかになった。これには、以前存在しなかった新しい学部の出現と、学部名称の変更(農学部や文学部が別の名称に変わった場合が多い)の両方の影響がある。

(4) 複数の大学による連合大学院研究科の出現

複数の大学により構成される連合大学院研究科(United Graduate School)が各処に設立されている。連合農学研究科が最も多いが、教育学、獣医学、小児発達学等の連合研究科もある。

2-2 全学的組織の拡大

全学的組織が大幅に増えている。これらを系統的に分類することは難しいが、①学部・研究科を横断した研究組織、②全国共同利用型研究施設、共同利用・共同研究拠点、③時限的研究制度・資金によるプロジェクト型組織、④産学協同、地域連携のための組織、⑤教員や学生の教育・研究・学内生活のための支援組織(保健衛生、環境保全、留学生対応、男女共同参画等)などがある。

これらの組織の階層付けの仕方も大学によって様々である。大学直属の独立組織、共通性のあるセンターをまとめた機構や本部組織、事務局組織の一環などがある。

2-3 頻繁な組織変更

各大学では新しい状況への対応のために、組織の新設、改組が頻繁に行われている。時限的な組織が多いこともあるが、学部や研究科の再編、その下の学科や専攻の組織替えも始終行われている。このため、組織構造や変遷の把握が一層困難になっている。

3 NISTEP 大学・公的機関名辞書を用いた論文データベースの機関同定

機関名辞書の適用例として、この辞書と論文データベースをリンクさせることによる日本の機関の研究活動データの分析がある。その際重要なのは、論文データベース中の著者所属機関データと機関名辞書中の機関(代表機関または下部組織)との対応付け(機関同定)である。ここでは、NISTEP で開発している機関名同定プログラムを Web of Science Core Collection (WoSCC)データベースに適用した結果(2015年5月実施)について述べる。

3-1 全体的な結果

機関同定結果の概要を概要図表 6 に示す。この表は、1996-2013年の18年間で6年ずつの3つの時期に区切り、機関同定されたデータと機関同定されなかったデータの出現頻度を示したものである。機関同定されたデータにはセクター別内訳を含む。機関同定されなかったデータには、セクターが推定される場合はセクター分類を与え、病院であることが推定される場合には病院フラグを付与するが、この図表にはそれらの出現度数も示している。なお、この表の数字は整数カウントによる出現機関数なので、1論文中に複数の出現機関データがあればそれぞれが1件として数えられる。1論文中に同一機関の異なる下部組織のデータが同定された場合も重複してカウントされる。

機関同定されたデータはどの時期もほぼ94%である。同定できなかった6%のデータには、機関名辞書に未収録の機関によるもの、機関名辞書に収録されている名称と異なる表記によるものがあり、これらのうちある程度出現頻度が高いものは辞書へのデータ追加を行っている。しかしこれ以外に、表記が不備あるいは曖昧で判断しづらい事例が相当数存在する。このようなデータは、無理に同定機関を決めると誤同定になる可能性があるため、むしろ同定不能のままにしておくことが望ましい。従って、機関同定率はせいぜい95%が限界ではないかと考えられる。

機関同定処理では、所属機関データと機関名辞書の名称データのマッチングを次の順で行う。

- ① 最長マッチング：所属機関データ表記単語列に完全に内包される機関名辞書単語列のうち、最長単語列のデータに相当する機関に同定する。
- ② 曖昧マッチング：①の同定ができなかった場合、N-gram³とレーベンシュタイン距離⁴を使用したマッチングを適用する。
- ③ 郵便番号マッチング：②の同定ができなかった場合に適用する。

このどの時点で同定がなされたかを示すのが概要図表 7 である。98%以上が最長マッチング処理で同定されており、この処理の重要性が判る。この表のマッチング種別コード C は、機械同定後の目視処理により同定を修正されたデータである。

³ 機関名辞書中のデータをN字(たとえばN=3)ごとに区切った文字列の索引を作り、所属機関データに含まれる文字列を高速検索する方法。

⁴ N-gram でマッチした候補データについて、挿入、欠失、置換を考慮した比較によりマッチングを行う方法。

概要図表 6 機関同定結果の概要

セクター		1996-2001	2002-2007	2008-2013	全期間
機 関 同 定	1 国立大学	351,325	440,229	485,315	1,276,869
	2 国立短大	132	90	2	224
	3 国立高専	2,755	3,193	3,136	9,084
	4 公立大学	33,287	42,334	45,443	121,064
	5 公立短大	443	281	123	847
	6 公立高専	204	310	279	793
	7 大学共同利用機関	9,802	13,546	13,861	37,209
	8 国の機関	35,275	15,106	12,322	62,703
	9 特殊法人・独立行政法人	36,795	102,854	124,479	264,128
	10 地方公共団体の機関	14,252	17,256	21,791	53,299
	12 私立大学	109,164	141,533	176,635	427,332
	13 私立短大	960	802	583	2,345
	14 私立高専	19	24	31	74
	15 会社	66,365	58,262	49,683	174,310
	16 非営利団体	16,995	20,792	21,799	59,586
	17 その他	167	214	187	568
	小計(A)		677,940	856,826	955,669
機 関 同 定 不 能	セクター同定	12,608	14,414	16,693	43,715
	病院同定	18,512	23,297	32,583	74,392
	同定不能	12,803	15,614	17,181	45,598
	小計(B)	43,923	53,325	66,457	163,705
総計(C=A+B)		721,863	910,151	1,022,126	2,654,140
機関同定率(%A/C)		93.9	94.1	93.5	93.8

概要図表 7 機関同定されたデータのマッチング種別分布

種別コード	マッチングの内容	同定数	比率(%)
L	最長マッチングによる同定	2,444,791	98.17
M	曖昧マッチングと郵便番号マッチングの結果が一致	898	0.04
U	曖昧マッチングによる同定	13,017	0.52
P	郵便番号マッチングによる同定	18,773	0.75
C	目視チェックにより同定を修正	12,956	0.52
合 計		2,490,435	100.00

※Cは目視チェックにより同定を修正したレコード

3-2 32 大学の下部組織同定

最長マッチング処理で 32 大学のいずれかが代表機関として同定されると、次の手順で下部組織同定がなされる。②と③は大学下部組織の同定精度を向上させるために開発した補助ツールである。

- ① 機関名辞書とのマッチング
- ② 拡張された補助セット⁵とのマッチング
- ③ 下部組織統計辞書⁶とのマッチング

⁵ WoSCC に一定回数出現する大学下部組織の表記のうち、機関名辞書とのマッチングがなされないものについて、目視で大学下部組織との対応付けを行ったデータ。

①～③のいずれともマッチしなければ、代表機関である大学に同定される。概要図表 8 に、各大学の下部組織同定率及び下部組織同定されたデータが①～③のいずれとマッチしたかの内訳を示す。

概要図表 8 下部組織マッチングがなされたツール別内訳(%)

大学	出現度数 (A)	うち下部組織 同定(B)	下部組織同定 率(B/A%)	部局同定のマッチング辞書内訳(Bを分母としたときの%)			
				①機関名辞書	②拡張補助セット	③下部組織統計辞書	②+③
北海道大学	59,987	55,082	91.8	89.1	8.1	2.8	10.9
東北大学	88,772	81,039	91.3	86.1	3.4	10.5	13.9
筑波大学	36,458	30,217	82.9	92.5	5.8	1.7	7.5
群馬大学	14,255	13,395	94.0	78.1	1.4	20.5	21.9
千葉大学	26,573	24,357	91.7	92.3	3.9	3.8	7.7
東京大学	150,391	140,647	93.5	78.7	3.2	18.2	21.3
東京医科歯科大学	18,374	8,055	43.8	71.8	0.0	28.2	28.2
東京工業大学	45,466	37,689	82.9	48.0	27.2	24.8	52.0
東京農工大学	11,279	8,538	75.7	44.9	35.6	19.5	55.1
新潟大学	20,333	18,458	90.8	85.8	3.8	10.4	14.2
富山大学	13,183	7,391	56.1	76.8	18.6	4.6	23.2
金沢大学	19,702	18,156	92.2	91.4	5.0	3.5	8.6
信州大学	16,271	15,308	94.1	86.7	2.4	10.9	13.3
岐阜大学	15,065	13,994	92.9	84.6	5.2	10.3	15.4
名古屋大学	60,642	53,841	88.8	77.0	8.4	14.5	23.0
京都大学	108,408	99,647	91.9	82.0	6.4	11.6	18.0
大阪大学	92,256	87,154	94.5	90.1	1.1	8.8	9.9
神戸大学	25,206	23,013	91.3	90.7	7.9	1.4	9.3
岡山大学	28,345	26,083	92.0	86.8	4.6	8.6	13.2
広島大学	33,641	28,367	84.3	66.8	4.7	28.5	33.2
徳島大学	15,025	13,266	88.3	87.1	2.1	10.8	12.9
九州大学	64,597	58,754	91.0	85.3	10.2	4.5	14.7
長崎大学	17,828	16,403	92.0	97.1	1.1	1.8	2.9
熊本大学	17,671	15,661	88.6	84.4	9.3	6.3	15.6
大阪市立大学	16,718	15,141	90.6	89.1	3.8	7.1	10.9
大阪府立大学	12,000	8,277	69.0	63.2	11.7	25.1	36.8
慶應義塾大学	29,248	26,531	90.7	80.8	0.7	18.5	19.2
東海大学	12,947	10,496	81.1	81.7	14.6	3.7	18.3
東京理科大学	15,277	13,523	88.5	63.5	18.5	18.0	36.5
日本大学	18,972	17,914	94.4	89.9	3.8	6.3	10.1
早稲田大学	17,361	13,838	79.7	65.8	26.5	7.7	34.2
近畿大学	12,243	11,088	90.6	83.7	12.0	4.3	16.3
全体	1,134,494	1,011,323	89.1	81.9	6.7	11.4	18.1

全大学での下部組織同定率は 89.1%で、19 大学が 90%を越えるが 80%未満が 5 大学ある(これについては後述)。また、下部組織同定に対する①、②、③の貢献率はそれぞれ 81.9%、6.7%、11.4%であり、補助ツールとして導入した②と③がかなりの効果を及ぼしたことが判る。②と③を合わせた貢献率は大学により 3%～55%と大きな差がある。この数字と下部組織同定率には相反関係がある(Pearson 相関係数は -0.46)ことから、これらの補助ツールはもともと下部組織同定率が低い大学でより大きな効果を発揮しており、大学間の下部組織同定率の格差縮小に寄与していると言える。

下部組織同定がされない(つまり代表機関である大学に同定される)理由は次の 2 つに大別される。

- (1) 所属機関データに大学名(もしくは大学名+キャンパス名)しか書かれていない。
- (2) 表記されている下部組織名が機関名辞書や補助ツールとマッチしない。

このことを概要図表 9 で説明する。ここでは、3 つの時期すべてで下部組織同定率が 85%以上の 22

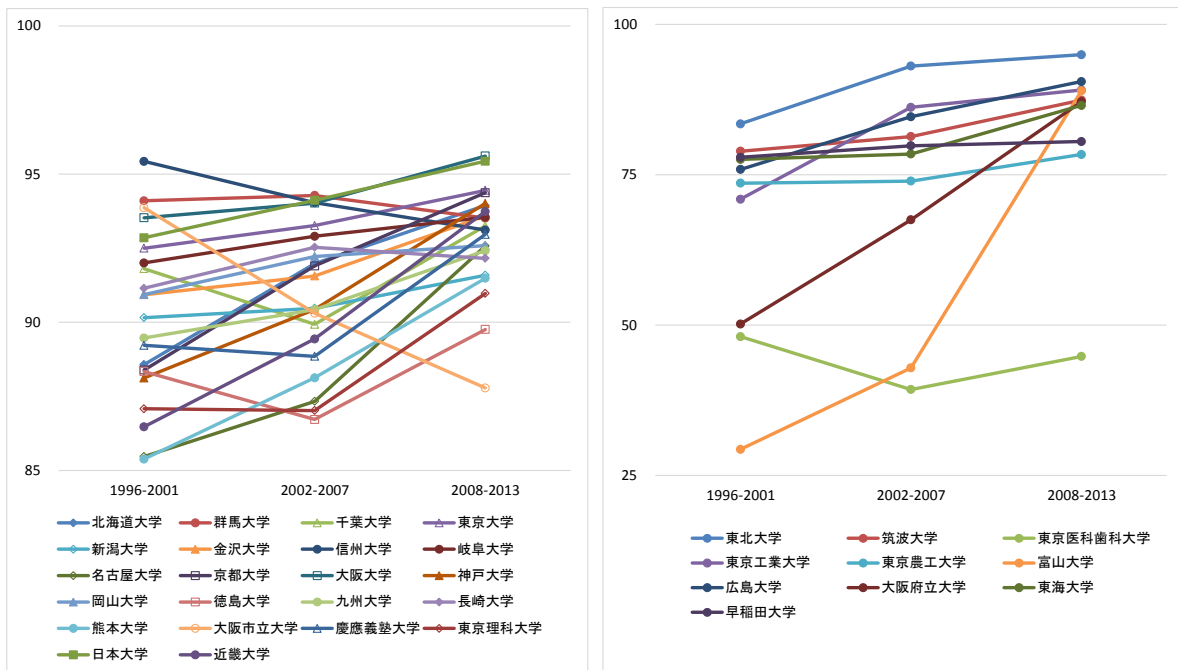
⁶ WoSCC に一定回数出現する大学下部組織の表記と大学下部組織との対応付けを、統計的に行ったデータ。

大学(概要図表 9(a))とそれ以外の 10 大学(概要図表 9(b))を別の図に示す。目視確認から、概要図表 9(a)に示した大学では、下部組織同定されなかったデータの多くが(1)に属することが分かっている。言い換えると、大学のデータの 5~15%はこの種の表記がなされている。これらのデータから下部組織同定をすることは不可能なので、現状では 90%前後が下部組織同定率の限界と考えられる。一方、概要図表 9(b)に示した大学では、下部組織が表記されているにもかかわらず、下部組織同定が行われていない場合も多く、(1)とともに(2)のデータが相当混じっていると考えられる。

概要図表 9 下部組織同定率の変遷

(a) 全時期で同定率 85%以上の 22 大学

(b) それ以外の 10 大学



東京医科歯科大学は、他の大学に比べて下部組織同定率がきわめて低く、どの時期も 50%に達しない。この大学は、学部では医学部と歯学部、大学院では医歯学総合研究科が規模において大部分を占めているため、所属機関表記で第 2 階層下部組織に対応する学部名、研究科名が省略されて、第 3 階層下部組織である学科名等が直接表記されることが非常に多い。学部名や研究科名が書かれたデータが少ないと、下部組織統計辞書を作るのに必要な統計的信頼性を持ったデータが得られないことになる。

東京医科歯科大学ほどではないが、東京農工大学、早稲田大学の下部組織同定率もあまり高くない。両大学とも、その大きな理由は、下部組織同定ができなかったデータのうち上記(1)のデータの割合が高いことである。

富山大学と大阪府立大学は、3 つの時期の間で下部組織同定率がそれぞれ 29.3%→42.9%→88.9%、50.2%→67.5%→87.1%と急上昇しているが、これには両大学において他大学との統合があったことが影響

している。ここでの集計にはこれら統合された大学のデータも含んでいるが、それらの統合前の大学の同定データは下部組織同定に数えられていない。

概要図表 9(b)にあるこれ以外の大学(東北大学、筑波大学、東京工業大学、広島大学、東海大学)は、いずれも下部組織同定率が徐々に上昇し、2008-2013年には87%以上に達している。東京医科歯科大学を除く31大学では、2008-2013年における下部組織同定率は75%を越えているので、各種分析に利用する際のベースラインとしての下部組織同定はなされていると考えられる。

4 まとめと今後の課題

本報告書では、NISTEPにおける機関名辞書の整備と、その活用例としての論文データベース中の著者所属機関同定について述べた。特に、32大学における下部組織情報の機関名辞書への収録と、その結果可能になった下部組織レベル同定に焦点を当てた。

以上の取り組みを踏まえ、機関名辞書の継続的維持と大学下部組織情報の拡張の2点が、今後の課題として重要と考えられる。

第1の課題に関しては、機関やその下部組織は常に変遷しており、その情報を確実にとらえてアーカイブ化するためには、不断の状況把握・情報更新が必要である。情報収集とデータ更新を行う体制と予算措置を継続的に講じられるよう検討を進めたい。それにより、大学、短大、高専、大学共同利用機関、国の機関、特殊法人・独立行政法人、学校法人に属する代表機関とその下部組織については毎年少なくとも1回、地方公共団体の機関、民間企業及び非営利団体については数年に1回程度を目標に、関係のWebサイト等の調査を行い、機関の変遷、英語名の変更等の有無を確認して必要な情報更新を行いたい。

第2の課題について、現在下部組織情報の網羅的収録の対象としている32大学は、1996～2010年の期間の論文数を基準として選定している。しかし、その実態は年とともに変化するので適宜見直しが必要である。本来は、ある程度の研究規模を持つ総合大学をすべて対象とするのが望ましい。この作業を効率的に行うためには、毎年確実に各大学の組織変遷情報を知る適当な方法を検討する必要がある。

本編

(裏白紙)

1 背景と目的

我が国における研究開発の動向を把握し、研究開発に関する方針・政策を検討するためには、以下の情報が基礎データとして不可欠であろう。

- 産・学・官それぞれにどのような研究機関が存在するか。
- それらの機関にどれだけの研究人材や研究資金が配分されているか(研究インプット)。
- それらの機関からどれだけの研究成果が出されているか(研究アウトプット)。

このようなデータは、国レベルの政策策定にも、個別の研究機関における研究戦略・研究計画の検討にも重要と考えられる。

この種のデータを取得するための情報源はいろいろ存在する。研究インプットに関しては、科学技術要覧等に収められる統計データの他に、調査・研究目的では科学技術研究調査などの個票データも利用できるようになった。また、最も主要な研究資金である科研費については KAKEN: 科学研究費助成事業データベース(<https://kaken.nii.ac.jp/>)が、研究者については researchmap (<http://researchmap.jp>) が公開されている。研究アウトプットに関しては、内外の論文データベースや特許データベースが重要な情報源である。

しかし、これらの情報源を用いて、ある観点から機関別にデータを整理・分析しようとする、多くの困難に直面することが多い。その主な原因には次のことが挙げられる。

- ① 機関の名称が一致しない。同じ機関に対して、情報源によって異なる名称が使われ、同じ情報源の中でも機関表記のゆれが存在する場合がある。当然のことながら、海外の情報源では英語機関名が用いられる。
- ② 大規模な研究機関では、機関レベルのみならず下部組織レベルでのデータ分析が必要な場合がある。しかし、それぞれの機関の組織構成の情報を一括して得ることは難しい。
- ③ 機関は統廃合や名称変更によりしばしば変遷する。また、機関内の組織変更も頻繁に行われる。ある期間にわたってデータ分析を行う場合には、この変遷情報を知ることが重要である。
- ④ データをセクターごとに集計したい場合があるが、多くの情報源には機関が属するセクターの情報が示されていないか、大学、公的機関、民間機関程度の粗い分類がされている程度である。

以上の問題点に対処する一つの方法として、研究機関に関する基本的情報を系統的に収録した辞書を整備することがある。科学技術・学術政策研究所(NISTEP)における「NISTEP 大学・公的機関名辞書」の作成、公開は、この目的に沿った一つの試みである。この辞書はまだ開発途上であるが、上記に挙げた問題点に対処できることを念頭に、和英の機関名称、機関内の階層関係、機関の変遷、所属セクター等の情報の充実に努めている。

NISTEP では、機関名辞書及び関連のデータの各方面での活用を期待して公開している[1]が、自身で行う調査・分析にもこの辞書を活用している。その一つが、論文データベース中の所属機関データの

同定(機関名寄せ)と機関別研究アウトプット分析である。これらについては参考文献[2]で報告しているが、その後、辞書の構成、同定の方法に種々の改善を行った。特に、発表論文の多い 32 の大学については、学部・研究科や附置研究所等の下部組織を網羅的に機関名辞書に収録し、下部組織レベルでの同定を可能とした。

本報告書では、参考文献[2]以降の改善に重点を置いて、以下の順に報告する。第 2 章では、機関名辞書の意義・特徴とともに、その構成、収録情報、情報収集法の概要を述べる。第 3 章は機関名辞書への大学下部組織情報の網羅的収録の対象とした 32 大学の選定と情報収集方法を示した後、最近の大学の組織構造がきわめて多様化・複雑化している状況とその情報を得ることの困難性を説明する。第 4 章は機関名辞書を用いた論文データベースの機関同定の事例報告で、データベース中の著者所属機関データの表記ゆれ等、同定を困難にする原因を整理した後、これらの困難性に対処するための機関名辞書の情報付加、補助ツールの開発、同定アルゴリズムの工夫などについて述べる。第 5 章では、Web of Science Core Collection データベースに対して得られた同定の結果を、32 大学の下部組織同定を中心に述べる。第 6 章は、下部組織同定をより容易にするため、今後考えられる方策について検討・提言を行う。

2 NISTEP 大学・公的機関名辞書の継続的整備

2-1 大学・公的機関に関する基本的情報のアーカイブ化

我が国における研究開発の動向を把握し、研究開発に関する方針・政策を国レベル、機関レベルで検討する上で、研究開発を行っている機関に関する情報は、最も基盤的なデータの一つである。特に、研究開発費の相当部分が国の予算により行われている大学及び公的機関(国の機関、独立行政法人等)に関する情報整備は、国の科学技術政策を推進する上で重要である。

しかし、これらの機関についての基本的情報(日本語と英語の正式名称と通称や略称、属するセクター、所在地、主な下部組織、変遷情報等)をまとめた資料はほとんどない。かつて隔年出版されていた「全国試験研究機関名鑑」(文部科学省監修、丸善出版)は、2008/2009 年版以降出版されていない。機関に関する検索サイトはいくつかあるが(科学技術振興機構のJ-Globalなど)、統合的、分析的利用には適しない。官庁のホームページには機関タイプ別のリストがあり(文部科学省の高等教育機関一覧、内閣府の独立行政法人一覧など)、Wikipedia にもいろいろな機関タイプのリストがあるが、下部組織等の詳しい情報は得られない。

機関には常に変遷が伴う。国立大学では 21 世紀に入ってから統合が進み、私立大学・短期大学の新設・廃止や短期大学から 4 年制大学への移行(あるいは 4 年制大学との統合)は数多い。国の機関や独立行政法人でも様々な動きがある。このような変遷情報が一元的にアーカイブ化されておれば、機関の活動を時系列で追跡するときの困難さが軽減されるであろう。

NISTEP では、研究開発を実施している我が国の機関(大学と公的機関に重点)の基本的情報を収録する「NISTEP 大学・公的機関名辞書」(以下「機関名辞書」という)の作成、維持、公開に取り組んでいる。機関名辞書の整備は、2011 年度から行われている文部科学省の「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」の一環としての「データ・情報基盤の構築」の一部であり、上記の認識のもとに、研究開発に関する機関の基本的な情報を継続的に取得し、アーカイブ化することを目標としている。

2-2 機関名辞書への下部組織レベル情報の付加の必要性

研究開発の動向を把握するとき、学部、大学院研究科、附置研究所等の下部組織レベルでの調査・分析が望ましい機関がある。大規模な総合大学では、いろいろな分野の研究が多数行われているので、詳細な研究活動の状況を把握するには、その下部組織(学部、大学院研究科、附置研究所等)ごとの情報が重要な場合がある。また、大学共同研究機関や一部の独立行政法人では、元々独立していた機関を「機構」という名の統合組織にまとめているが、これらについても、下部組織ごとの分析が意味を持つ場合が多い。このようなことから、機関名辞書では重要な下部組織も収録している。

後述するように、大学の下部組織の情報収集には大きな困難が伴う(情報が探しにくい、変遷が頻繁で現存しない下部組織の情報は Web サイト等から消えてしまう等)。また、大学により下部組織の構成や

名称の付け方は様々で、統一的体系をとることが望ましい「辞書」の構造への合わせ方にも問題が生ずる。大学の下部組織レベルでのデータ整理や分析が、必要性は高いのにあまり行われていない理由の一つは、このような基本情報の収集・整理の難しさにあると考えられる。

しかし、このような困難性があるからこそ、どこかで一元化された情報がアーカイブ化されていないと、必要な調査や分析が行われないことになる。NISTEP の機関名辞書は、大学下部組織情報のアーカイブ化を一つの重要な目標としている。そのためには、変遷する下部組織の情報(どの学部・研究科等が、いつ、どの学部・研究科等が変わったか)を不断に蓄積する必要がある。

2-3 機関名辞書の概要

以下に述べる機関名辞書についての記述は、最新の公開バージョンである「NISTEP 大学・公的機関名辞書(Version 2015.1)」の内容によるものである。

機関名辞書には、研究開発を行っている日本の機関(以下「代表機関」という)とその主な下部組織の名称(日本語と英語)、所属するセクター、階層関係、変遷等の情報を収録する⁷。その概要については、参考文献[2]の第3章に述べているが、その後その収録基準等について更に検討し、2015年度はじめに、その方針に基づいて再構成を行った。その結果、含まれる情報をコアセットと補助セットの2つのデータセットに分けて収録することとした。図表1にコアセットと補助セットの関係性を示す。

図表 1 機関名辞書のコアセットと補助セット

コアセット

- 網羅的なリストの継続的な収集・アーカイブ化
- データ・情報基盤として公表

補助セット

- 論文データベース(Web of Science, Scopus)等の機関名同定に必要な補助的な情報の蓄積
- NISTEP内の分析で利用(機関名辞書を用いて論文データベースの名寄せを行うプログラムも併せて開発)
- 分析結果や分析の過程で得られたデータ(大学・公的機関名英語表記ゆれテーブル等)を公表

コアセットは機関名辞書の基本情報を収録し、NISTEP の「データ・情報基盤」の一つとして公開される。

⁷ 機関名辞書では、以下のような機関は下部組織ではなく代表機関としている。

(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：国立高専は、2004年以降、独立行政法人国立高等専門学校機構の下の組織になっているが、機関名辞書では個々の高専を代表機関としている。

(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(気象庁高層気象台、厚生労働省の検疫所等)は、属する省庁の下部組織とする。

(d) 地方公共団体の公設試験研究機関等

補助セットには、コアセットへの収録基準に合致しないが、論文データベース中の所属機関データの同定や分析の作業上収録の必要性が高い機関や情報を収録する。

ここでは、公開対象とするコアセットに収録する機関と情報を示し、補助セットについては、機関・下部組織同定について記述した 4-4 の中で触れる。

2-3-1 コアセットに含まれる代表機関

コアセットに含まれる代表機関は、つぎに示す条件に合う機関である。

- ① 現存する全ての大学、短期大学、高等専門学校、大学共同利用機関及び学校法人
- ② 研究開発を行っている国の機関、特殊法人・独立行政法人等、地方公共団体の機関、非営利法人、その他の機関(日本所在の国際機関等)
- ③ ①または②のセクターに属し、この 15 年ほどの間に統廃合、名称変更等があつて現存しない機関のうち主要なもの

民間企業はコアセットでなく補助セットに収録する。その理由は、別途に構築、公開している NISTEP 企業名辞書に、我が国の企業に関する主要な情報が収録されているためである。しかし、論文データベースには企業から発表される論文も多数含まれているので、補助セットを用いてそれらの同定、分析を可能としている。

2-3-2 コアセットに含まれる下部組織

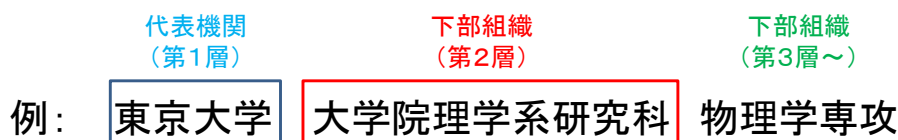
上記の代表機関に属する以下の下部組織を収録する。現存しない下部組織も含む。原則としては第 2 階層下部組織(代表機関の直下の下部組織)までを対象とする。図表 2 に代表機関-第 2 階層下部組織-第 3 階層下部組織の関係性のイメージを示した。

- ① 研究論文数の特に多い大学の下部組織。当面 32 大学(3-1 参照)を対象とし、それらの第 2 階層下部組織を網羅的に収録する(事務組織を除く)。病院及び共同利用・共同研究拠点となっている下部組織については第 3 階層下部組織(医学部の附属病院など)であっても収録対象とする。第 3 章では、この 32 大学の下部組織情報の収集に焦点を当てる。
- ② その他の大学に属する共同利用・共同研究拠点となっている下部組織。
- ③ 大学共同利用機関である 4 つの機構に直属する研究所等。
- ④ 国の機関及び特殊法人・独立行政法人等に所属する病院及び大学校。
- ⑤ 独立行政法人宇宙航空研究開発機構、独立行政法人国立博物館、独立行政法人国立文化財機構、独立行政法人水産総合研究センター、独立行政法人農業・食品産業技術総合研究機構に直属する研究所等⁸。

⁸ ここに挙げる独立行政法人の中には、2015 年 4 月 1 日に国立研究開発法人に移行したものが含まれているが、本報告書の記述は

- ⑥ 独立行政法人科学技術振興機構の主要プロジェクト事業(CRESTO、ERATO 等)。
- ⑦ 国の機関及び特殊法人・独立行政法人等に直属するその他の下部組織で、研究開発活動の分析の観点から重要なもの(独立行政法人理化学研究所脳科学総合研究センター、独立行政法人日本貿易振興機構アジア経済研究所等)。

図表 2 代表機関-第 2 階層下部組織-第 3 階層下部組織



2-3-3 コアセットに含まれる情報

機関名辞書(コアセット)には、上記に該当する代表機関及び下部組織に対する以下の情報を収録する。

- ① NISTEP が与える機関 ID (代表機関、下部組織のそれぞれに独立して与える)
- ② 日本語と英語の正式名
- ③ 所属するセクター(17 に分類、図表 3 を参照)
- ④ 下部組織とそれが属する代表機関とのリンク
- ⑤ 現存しない代表機関、下部組織の変遷情報(移行の種別、移行年月日、継承先がある場合その機関)
- ⑥ 郵便番号

これらのうち、日本語正式名と所属セクターは、すべての代表機関、下部組織に付与されている。英語正式名が付与されているのは全機関(下部組織を含む)のうち 46.0%であるが、大学と公的機関(図表 3 のセクター番号 1～9 及び 12～14)に限れば 96.8%に付与されている。また、郵便番号は、代表機関の約 30%に少なくとも 1 個付与されている。

機関名辞書の最新版(Version 2015.1)のコアセットに収録されているセクター別の機関数を図表 3 に示す。

機関名辞書の最新公開バージョン(Version 2015.1)に基づいており、それは 2015 年 1～2 月時点の調査によるものである(図表 5 参照)ため、ここではすべて独立行政法人とする。

図表 3 機関名辞書(コアセット)の収録機関数 (Version 2015.1)

セクター	セクター	現存機関			非現存機関			合計		
		代表機関	下部組織	小計	代表機関	下部組織	小計	代表機関	下部組織	小計
1	国立大学	86	1,268	1,354	15	122	137	101	1,390	1,491
2	国立短大				26		26	26		26
3	国立高専	51		51	8		8	59		59
4	公立大学	85	60	145	16	1	17	101	61	162
5	公立短大	17		17	45		45	62		62
6	公立高専	3		3	4		4	7		7
7	大学共同利用機関	4	24	28	3		3	7	24	31
8	国の機関	40	9	49	57	7	64	97	16	113
9	特殊法人・独立行政法人	88	224	312	37	29	66	125	253	378
10	地方公共団体の機関	685		685	90		90	775	0	775
11	学校法人	663		663	1		1	664		664
12	私立大学	612	364	976	41	17	58	653	381	1,034
13	私立短大	341		341	223		223	564		564
14	私立高専	3		3				3		3
15	会社									
16	非営利団体	3,568		3,568	318		318	3,886		3,886
17	その他	5		5				5		5
	総計	6,251	1,949	8,200	884	176	1,060	7,135	2,125	9,260

2-4 機関名辞書に収録するデータの収集

2-4-1 大学と国に関する機関

(1) 代表機関の情報

大学、短大、高専、大学共同利用機関、国の機関、特殊法人・独立行政法人、学校法人の各セクター（これらは、図表 3 のセクター番号 1～9 及び 11～14 に当たる）に属する代表機関については、毎年少なくとも 1 回、図表 4 に示す Web サイトのリストを調べて存在を確認する（なお、図表 4 に示した情報は、2015 年 7 月時点の情報である。これらのページが維持・管理されなくなった場合は、再度、適当な Web サイトを設定する）。新しくリストに記載された機関は辞書に登録する。リストからなくなった機関については、Web 検索により、変遷の区分（統合、廃止、名称変更の 3 種に分類）、変遷のあった年月日（月日まで判らない場合は年のみ）及び継承機関を調査して、その情報を辞書に追加する。統合及び名称変更の場合は原則として継承機関が存在するが、廃止の場合も業務を受け継いだ機関が判ればそれを継承機関とする。継承機関が辞書に未登録の場合は、新しい機関として登録する。代表機関であったものが統合によって別の代表機関の下部組織になったときは、その下部組織を辞書に登録してこれを継承機関にする場合と、下部組織は登録せずに統合先代表機関を継承機関とする場合とがある。

辞書に新たに登録された機関については、その機関の Web サイトから、正式名称（日本語と英語）と郵便番号の情報を入手する。機関 Web サイトから英語名称が判らない場合は Wikipedia 等から入手することもある（この場合は辞書では正式名称とはせず別名扱いとする）。

図表 4 機関の存在確認のための Web サイト(2015 年 7 月現在)

機関の種類	Webサイト名	URL
高等教育機関(大学、短大、高専)	文部科学省関係リンク集>教育	http://www.mext.go.jp/b_menu/link/1294885.htm
大学共同利用機関	文部科学省関係リンク集>科学技術・学術>大学共同利用機関法人	http://www.mext.go.jp/b_menu/link/daiyou.htm
国の行政機関の組織	国の行政機関の組織の一覧表	http://www.cas.go.jp/jp/gaiyou/jimu/jinjikyoku/satei_01_03_01.html
国の「施設等機関」	Wikipedia「施設等機関」	http://ja.wikipedia.org/wiki/%E6%96%BD%E8%A8%AD%E7%AD%89%E6%A9%9F%E9%96%A2
国の「特別の機関」	Wikipedia「特別の機関」	http://ja.wikipedia.org/wiki/%E7%89%B9%E5%88%A5%E3%81%AE%E6%A9%9F%E9%96%A2
独立行政法人	内閣府・独立行政法人一覧	http://www.cas.go.jp/jp/gaiyou/jimu/jinjikyoku/files/satei_01_05_21.pdf
独立行政法人	Wikipedia「日本の独立行政法人一覧」	http://ja.wikipedia.org/wiki/%E6%97%A5%E6%9C%AC%E3%81%AE%E7%8B%AC%E7%AB%8B%E8%A1%8C%E6%94%BF%E6%B3%95%E4%BA%BA%E4%B8%80%E8%A6%A7
国立研究開発法人	Wikipedia「国立研究開発法人」	https://ja.wikipedia.org/wiki/%E5%9B%BD%E7%AB%8B%E7%A0%94%E7%A9%B6%E9%96%8B%E7%99%BA%E6%B3%95%E4%BA%BA
特殊法人	内閣府・特殊法人一覧	http://www.cas.go.jp/jp/gaiyou/jimu/jinjikyoku/files/satei_01_05_22.pdf
認可法人	Wikipedia「認可機関」	http://ja.wikipedia.org/wiki/%E8%AA%8D%E5%8F%AF%E6%B3%95%E4%BA%BA
学校法人	日本私立学校振興・共済事業団・学校法人情報検索システム	http://meibo.shigaku.go.jp/

(2) 下部組織の情報

辞書に下部組織情報を網羅的に収録する32大学の下部組織情報収集については別途3-2で述べるので、ここではそれ以外の代表機関に属する下部組織の情報収集について記す。

機関名辞書のコアセットに含まれる下部組織(2-3参照)については、少なくとも年1回、それぞれが属する代表機関のWebサイトで存否を確認する。存在しなくなった下部組織については、代表機関の場合と同様、変遷区分、変遷年月日、継承組織の情報を追加する。新しい下部組織を登録する場合には、その英語名も調査する。

現在公開している機関名辞書(Version 2015.1)の情報の取得時期及び非現存機関のカバー範囲を図

表 5 に示す。従って、この取得時期以降の機関、組織の新設や変遷の情報は含まれていない。

図表 5 機関名辞書 (Version 2015.1) の情報の取得時期及び非現存機関のカバー範囲

機関の種別	最新情報取得時期	非現存機関のカバー範囲*
大学等(短大、高専、大学共同利用機関を含む)	2015年1月	2008年度以降
32大学の下部組織	2015年2月	2013年度以降
国の機関、特殊法人・独立行政法人等	2015年1月	2008年度以降

* 確実にカバーしている範囲であり、それ以前に現存しなくなった機関も含まれる。

2-4-2 地方公共団体の機関、民間企業、非営利法人に属する機関

機関名辞書の作成を開始した 2011 年に、これらのセクター(図表 3 のセクター番号 10 及び 15~17 に当たる)に属する機関は主に「全国試験研究機関名鑑」(2008/2009 年版)[3]に基づいて選択した。これらの機関については図表 4 に示すような適当なリストがないため、その後定期的な点検は行っておらず、気が付いた時点で適宜、機関の新登録、変遷情報の追加、英語名の追加や変更を行う程度であった。但し、地方公共団体の機関のうち地方独立行政法人については Web 上にリストが存在するので、これを用いて調査を行っている。また、主要な非営利団体(論文発表数が多い機関)については、財団・社団法人から公益あるいは一般財団・社団法人への移行の状況を 2014 年度に一斉調査し、機関名辞書の更新を行った。民間企業については、上述の NISTEP 企業名辞書と連携し、今後、必要な情報(特に変遷情報と所在地情報)の更新を行うこととしている。

地方公共団体の機関、民間企業、非営利法人に属する機関についても現状をより正確に把握するため、少なくとも数年に 1 回程度調査とデータ更新を行う必要がある。

2-4-1、2-4-2 に示した方法以外にも、変更の情報が得られればその都度辞書の修正を行っているが、その中で大きな情報源は、第 4 章で述べる論文データベースの機関同定である。このとき、同定できなかった所属機関データによく出現する表記から、機関名辞書に未登録の機関や下部組織、未収録の英語別名や機関略称の情報が得られる。

3 NISTEP 大学・公的機関名辞書への大学下部組織情報の付加

第2章で触れたように、機関名辞書には、32の大学の第2階層までの下部組織の情報を網羅的に収録することとしている。この章では、この32大学の選定と下部組織情報の収集の方法を述べた後、情報収集を難しくしている2つの問題、大学の組織構造の複雑化と、組織情報を示す大学Webサイトの表現の多様性について論ずる。

3-1 下部組織情報を網羅的に収録する大学とその選定方法

機関名辞書に下部組織情報を網羅的に収録する32大学と、収録されている下部組織数(機関名辞書の最新版(Version 2015.1)で現存している下部組織のみ)を図表6に示す。

これらの大学の選定は、Scopusの2011年版(2010年までのデータを収録)を用いて、次の条件を満たす大学とした。ここで延べ出現数とは、所属機関データに当該大学(下部組織を含む)が出現した延べ回数である。整数カウントでは、各所属機関データが1と数えられるので、1論文に同じ大学の異なる下部組織の所属機関データが複数存在すればその回数だけカウントされる。一方分数カウントでは、1論文中にn個の所属機関データがあればそれぞれに $1/n$ ずつが配分されるので、その総数は論文総数と等しくなる。なお、この計数はScopusのすべての分野、記事種類を対象としている。

(a) 次の2つの条件をいずれも満たす大学

- 1996-2010年の延べ出現数が、整数カウントで10,000以上、分数カウントで5,000以上
- 2006-2010年の延べ出現数が、整数カウントで4,000以上、分数カウントで2,000以上

(b) 上記の他、2006-10年の延べ出現数が整数カウントで6,000以上、分数カウントで3,000以上の大学

図表 6 機関名辞書に下部組織を網羅的に登録する 32 大学

種別	大学	下部組織数*	備考**
国立大学	北海道大学	86	○
	東北大学	79	○
	筑波大学	67	○
	群馬大学	21	
	千葉大学	48	○
	東京大学	74	○
	東京医科歯科大学	26	
	東京工業大学	67	○
	東京農工大学	24	
	新潟大学	27	
	富山大学	36	
	金沢大学	50	
	信州大学	30	
	岐阜大学	27	
	名古屋大学	67	○
	京都大学	70	○
	大阪大学	64	○
	神戸大学	42	○
	岡山大学	55	
	広島大学	53	○
徳島大学	39		
九州大学	134	○	
	長崎大学	35	
	熊本大学	38	
公立大学	大阪市立大学	35	
	大阪府立大学	24	
私立大学	慶應義塾大学	53	
	東海大学	71	
	東京理科大学	31	
	日本大学	51	
	早稲田大学	90	
	近畿大学	56	

* 機関名辞書収録下部組織数 (Version 2015.1に収録の現存する組織)

** ○印は2011年度から、それ以外は2012年度から収録。

3-2 下部組織情報の収集の実際

図表 6 に示す 32 大学中、○印を付けた 12 大学については 2011 年度から、その他の 20 大学については 2012 年度から下部組織調査を行い、その変遷情報を含めて機関名辞書に収録している。機関名辞書に収録する情報の一般的調査方法については 2-4 に述べたが、ここでは特に、これら 32 大学の下部組織に関する情報収集について述べる。

調査は、主に各大学の Web サイトにより行っているが、後述する事情(3-3、3-4 参照)により、収録すべき下部組織の漏れがあったり、第 3 階層以下の下部組織が混在していたりした。そこで、2014 年度に 32 大学の下部組織構成を詳細に調査し、機関名辞書への下部組織情報収録の方針を以下のように定め、可能な限りこの方針に従って収録を行うこととした。

- ① 各大学の下部組織情報を入手するページを規定する。2014 年度に用いたこれらのページの URL を図表 7 に示す。多くは、大学概要のサイトにある組織のページである(各大学 URL からの情報の取得は、2015 年 2 月に実施)。
- ② このページの記載に基づき、原則として第 2 階層にあたる下部組織(大学の直下の組織)を機関名辞書の収録対象とする。但し、以下の例外がある。
 - ・ 事務組織と見なされる下部組織は対象としない。
 - ・ 第 3 階層以下であっても病院(医学部附属病院等)及び共同利用・共同研究拠点は収録対象とする。ただし、動物病院は第 3 階層以下の場合は収録対象としない。
 - ・ 3-3 で述べる「第 1.5 階層」と見なせる下部組織が存在する場合は、その下の第 2 階層下部組織までを収録対象とする(現在これに該当する第 1.5 階層下部組織は、金沢大学の学域と早稲田大学の大学院の 2 種である)。
 - ・ 第 2 階層か第 3 階層か判断が難しいものは収録対象とする。
- ③ 第 4 章で述べる論文データベース中の所属機関データの下部組織レベル同定の便宜上、第 3 階層下部組織の情報収集を例外的に行うことがある。これらの情報は、機関名辞書の「補助セット」に収録される。
- ④ 下部組織の英語名が①に示すページから得られない場合は、その下部組織のサイト等から入手する。
- ⑤ 既存の機関名辞書にある下部組織が①のページに存在しない場合は、統廃合や名称変更があった可能性が高いので、大学の沿革情報のページやその他の情報源を調査する。

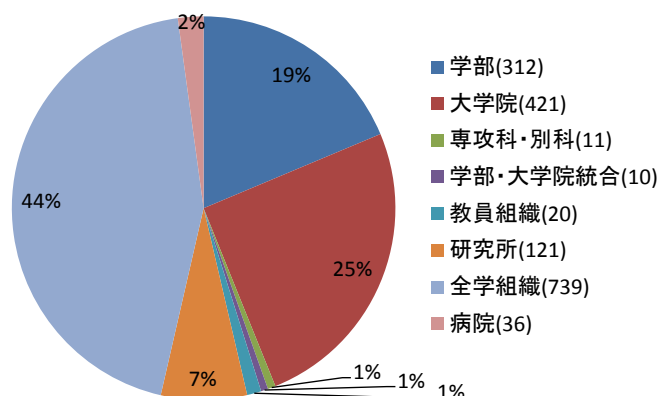
これらの大学下部組織は、図表 8 に示したように 8 種類に分類して機関名辞書に収録している(括弧内の数字は、機関名辞書(Version 2015.1)のコアセットに含まれる現存下部組織数)。

図表 7 32 大学の下部組織情報を入手する Web ページ(2015 年 2 月現在)

大学	URL(日本語1)	URL(英語1)
北海道大学	http://www.hokudai.ac.jp/introduction/dept/	http://www.oia.hokudai.ac.jp/about/university-profile/general-organization/
東北大学	http://www.tohoku.ac.jp/japanese/profile/about/05/about0502/	http://www.tohoku.ac.jp/en/about/organization_chart.html
筑波大学	http://www.tsukuba.ac.jp/about/organization.html	http://www.tsukuba.ac.jp/english/about/organization.html
群馬大学	http://www.gunma-u.ac.jp/html_hyouka/aboutus_3.html	http://www.gunma-u.ac.jp/english/html/aboutus_005.html
千葉大学	http://www.chiba-u.ac.jp/general/management/organization/documents/soshikizu.pdf	http://www.chiba-u.ac.jp/e/about/organization/index.html
東京大学	http://www.u-tokyo.ac.jp/gen01/b02_01_j.html	http://www.u-tokyo.ac.jp/en/about/organization/chart.html
東京医科歯科大学	http://www.tmd.ac.jp/outline/organization_rules/organization/index.html	http://www.tmd.ac.jp/english/outline/outline/organization/index.html
東京工業大学	http://www.titech.ac.jp/about/organization/index.html	http://www.titech.ac.jp/english/about/organization/index.html
東京農工大学	http://www.tuat.ac.jp/outline/kikouzu/upimg/20141001150427129889522.pdf	http://www.tuat.ac.jp/en/outline/kikouzu/upimg/201303231707001044453741.pdf
新潟大学	http://www.niigata-u.ac.jp/profile1/50_outline_020.html	http://www.niigata-u.ac.jp/e/about/10_about_030.html
富山大学	http://www.u-toyama.ac.jp/outline/about/pdf/h26/05.pdf#search=%E5%AF%8C%E5%B1%B1%E5%A4%A7%E5%AD%A6+%E7%B5%84%E7%B9%94	http://www.u-toyama.ac.jp/outline/about/pdf/h26/05.pdf#search=%E5%AF%8C%E5%B1%B1%E5%A4%A7%E5%AD%A6+%E7%B5%84%E7%B9%94
金沢大学	http://www.kanazawa-u.ac.jp/university/organization/index.html	http://www.kanazawa-u.ac.jp/e/academics/index.html
信州大学	http://www.shinshu-u.ac.jp/guidance/organization/structure/	http://www.shinshu-u.ac.jp/guidance/publication/summary/2013/gaiyou_en/#page=37
岐阜大学	http://www.gifu-u.ac.jp/about/overview/organization.html	http://www.gifu-u.ac.jp/en/en_overview_p6-p37.pdf
名古屋大学	http://www.nagoya-u.ac.jp/about-nu/org/	http://en.nagoya-u.ac.jp/about_nu/admin/sch/index.html
京都大学	http://www.kyoto-u.ac.jp/ja/about/organization/head/charts	http://www.kyoto-u.ac.jp/en/about/profile/faculty
大阪大学	http://www.osaka-u.ac.jp/ja/guide/organization	http://www.osaka-u.ac.jp/en/guide/organization/organization.html
神戸大学	http://www.kobe-u.ac.jp/info/organization/index.html	http://www.kobe-u.ac.jp/en/about-kobe-university/organization/organization-chart.html
岡山大学	http://www.okayama-u.ac.jp/tp/profile/profile03.html	http://www.okayama-u.ac.jp/en/tp/faculty/index.html
広島大学	http://www.hiroshima-u.ac.jp/top/gakubu_daigakuin/ http://www.hiroshima-u.ac.jp/top/kenkyusyo/	http://www.hiroshima-u.ac.jp/en/top/intro/
徳島大学	http://www.tokushima-u.ac.jp/about/profile/mechanism.html	http://www.tokushima-u.ac.jp/english/about/organization.html
九州大学	http://www.kyushu-u.ac.jp/university/organization/total_organization.php	http://www.kyushu-u.ac.jp/english/university/organization/organization.php
長崎大学	http://www.nagasaki-u.ac.jp/ja/about/guidance/outline/file/j06.pdf	http://www.nagasaki-u.ac.jp/en/about/organization/pdf/h24.pdf
熊本大学	http://www.kumamoto-u.ac.jp/daigakujuhou/kihonjoho/kyoiku_soshiki	http://ewww.kumamoto-u.ac.jp/about/organization/organization.html
大阪市立大学	http://www.osaka-cu.ac.jp/ja/about/university/organization_chart	http://www.osaka-cu.ac.jp/en/about/organization
大阪府立大学	http://www.osakafu-u.ac.jp/info/outline/org.html	http://www.osakafu-u.ac.jp/english/academics/index.html
慶應義塾大学	http://www.keio.ac.jp/ja/about_keio/organization/chart/kr7a4300000000ftp-att/20141101.pdf	http://www.keio.ac.jp/en/academics/Undergraduate.html
東海大学	http://www.u-tokai.ac.jp/about/organization/index.html	http://www.u-tokai.ac.jp/international/about/organization.html
東京理科大学	http://www.tus.ac.jp/info/foundation/pdf/sosiki20141001_02.pdf	http://www.tus.ac.jp/en/info/pdf/organization.pdf
日本大学	http://www.nihon-u.ac.jp/about_nu/board/chart/ http://www.nihon-u.ac.jp/research/institute/ http://www.nihon-u.ac.jp/affiliate_institute/	http://www.nihon-u.ac.jp/en/
早稲田大学	http://www.waseda.jp/top/assets/uploads/2014/04/organization.pdf	http://www.waseda.jp/top/assets/uploads/2014/10/organization.pdf
近畿大学	http://www.kindai.ac.jp/about-kindai/overview/download-data/organazation/organizational-office01.pdf	http://www.kindai.ac.jp/english/academics/index.html

* 機関名辞書の最新公開版 (Version 2015.1) のデータ調査を行った2015年2月現在のURL。その後更新されているページがある。

図表 8 32 大学の下部組織種別内訳(現存組織のみ)



3-3 最近の大学組織構造の傾向

2000年以前の大学の教員組織は、学部と大学院研究科、それに附置研究所が中心であった。学部の下には学科、学部附置研究所、学部附属施設(病院等)があり、また大学院研究科の下には専攻があり、それらの名称の付け方は、大学間でかなり共通性があった。もちろん、現在でもこの構造は保持されている場合もあるが、かなり大きな変動があり、大学の組織構成が多様化・複雑化している。このため、上述した「第2階層までの下部組織を網羅的に収録」の方針に従って下部組織を選別することは必ずしも容易でない。これについて、具体例を示しながら概観する。

3-3-1 学部、大学院の構成の多様化・複雑化

(1) 「学部」、「研究科」を表す呼称の多様化と教員組織の設置

2000年くらいまで、どの大学でも4年制は学部、大学院は研究科という呼称であった。現在でもこの呼称が多いが、学部の代わりに学類や学域等の呼称を、研究科の代わりに院、研究院、学府等の呼称を用いる大学が出現している。英語についても、以前は、学部は Faculty、研究科は Graduate School とほぼ決まっていたが、それ以外の呼称が増えている。

また、以前は、教員は学部に所属するのが普通であったが、大学院重点化以降は研究科所属が主流となった。更に最近では、学部や研究科以外に教員組織を置き、教員はそこに属して学部や大学院の教育を担当するという形にする大学が出現している(北海道大学、筑波大学、金沢大学、信州大学等)。その呼称は「系」、「研究院」、「研究域」など様々である。

図表 9 には、32大学の学部、大学院研究科及び教員組織の日本語と英語での呼称を示す(機関名辞書の最新版(Version 2015.1)を用いて集計)。呼称は大学により様々であり、たとえば“Faculty”は学部、大学院研究科、教員組織のどれにも使われている。また、同じ大学の中で、学部によって英語の呼称が異なる(多くは Faculty と School の2本立て)場合が32大学中14に達する。

図表 9 32 大学の学部、大学院研究科及び教員組織の呼称(2015年2月現在)

大学	種別	学部		大学院			教員組織*		学部・研究科統合機構	
		日本語名	英語名	日本語名	英語名	一部の研究科が教育組織と研究組織に分離	日本語名	英語名	日本語名	英語名
北海道大学	国立	学部	School (文・薬はFaculty)	研究科と院が混在	Graduate School		研究院	Faculty(またはResearch Faculty)		
東北大学	国立	学部	Faculty (医・歯・工・法はSchool)	研究科	Graduate School	教育情報学教育部(Education Division) 教育情報学研究部(Research Division)				
筑波大学	国立	学群	School	研究科	Graduate School		系	Faculty		
群馬大学	国立	学部	Faculty (理工のみSchool)	研究科(理工のみ学府)	Graduate School					
千葉大学	国立	学部	Faculty (医系はSchool)	研究科(医系は学府または研究院)	Graduate School					
東京大学	国立	学部	Faculty(教養のみCollege)	研究科(学際情報のみ学府)	Graduate School	学際情報学府(Graduate School)と情報学環(Interfaculty Initiative)				
東京医科歯科大学	国立	学部	Faculty	研究科	Graduate School					
東京工業大学	国立	学部	School	研究科	Graduate School					
東京農工大学	国立	学部	Faculty	学府	Graduate School		研究院	Institute		
新潟大学	国立	学部	Faculty	研究科	Graduate School					
富山大学	国立	学部	Faculty	研究科	Graduate School	理系は、 …学教育部(Graduate School … for Education) …学研究部(Graduate School … for Research)				
金沢大学	国立	【広域】学域 【狭域】学類	【広域】College 【狭域】School	研究科	Graduate School		【広域】研究域 【狭域】系	【広域】Institute 【狭域】Faculty		
信州大学	国立	学部	Faculty (医のみSchool)	研究科	Graduate SchoolとDivisionが混在		学域	School		
岐阜大学	国立	学部	Faculty (医系はSchool)	研究科	Graduate School					
名古屋大学	国立	学部	School	研究科	Graduate School					

* 灰色着色の教員組織は、大学院の研究組織に位置づけられている。

図表 9 32 大学の学部、大学院研究科及び教員組織の呼称(2015年2月現在)(続き)

大学	種別	学部		大学院			教員組織*		学部・研究科統合機構	
		日本語名	英語名	日本語名	英語名	一部の研究科が教育組織と研究組織に分離	日本語名	英語名	日本語名	英語名
京都大学	国立	学部	Faculty	研究科	Graduate School					
大阪大学	国立	学部	School (医系はFaculty)	研究科	Graduate School					
神戸大学	国立	学部	Faculty (医、経営はSchool)	研究科	Graduate School					
岡山大学	国立	学部	Faculty (医系はSchool)	研究科	Graduate School					
広島大学	国立	学部	Faculty	研究科	Graduate School	医歯薬保健学研究科(Graduate School)と医歯薬保健学研究院(Institute)				
徳島大学	国立	学部	Faculty	教育部	Graduate School		研究部	Institute		
九州大学	国立	学部	School	学府	Graduate School		研究院	Faculty		
長崎大学	国立	学部	Faculty (医系、工はSchool)	研究科	Graduate School					
熊本大学	国立	学部	Faculty (医系はSchool)	研究科	Graduate School	医系は、 …学教育部(Graduate School) …学研究部(Faculty)				
大阪府立大学	公立	学域(学部から移行中)	College (Schoolから移行中)	研究科	Graduate School		学群	不明		
大阪市立大学	公立	学部	Faculty	研究科	Graduate School					
慶應義塾大学	私立	学部	Faculty	研究科	Graduate School					
東海大学	私立	学部	School	研究科	School					
東京理科大学	私立	学部	Faculty (経営のみSchool)	研究科	Graduate School					
日本大学	私立	学部	College (医系はSchool)	研究科	Graduate School					
早稲田大学	私立	学部	School	研究科	Graduate School				学術院	Faculty
近畿大学	私立	学部	Faculty	研究科	Graduate School					

* 灰色着色の教員組織は、大学院の研究組織に位置づけられている。

(2) 階層関係の複雑化

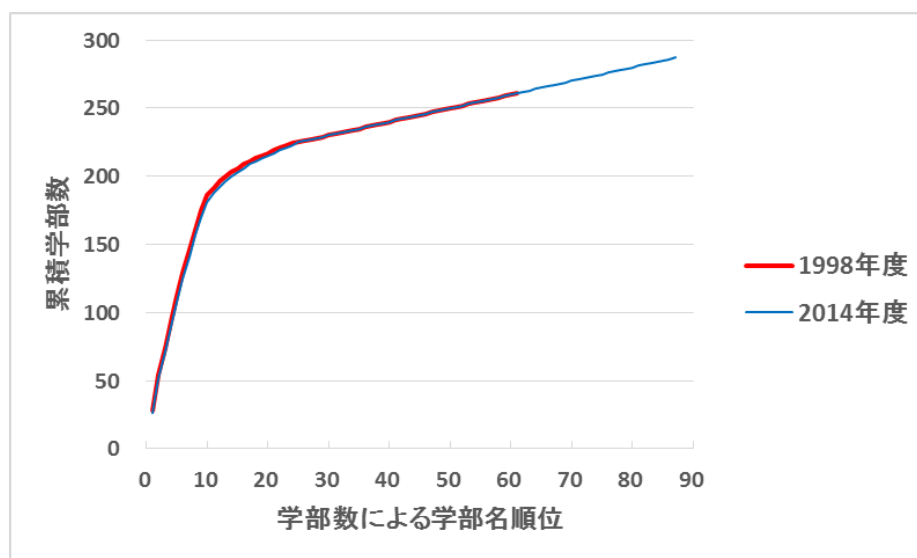
たとえば、金沢大学は2008年に、8つの学部を3学域(人間社会、理工、医薬保健)、16学類に再編した。16の学類は従来の学部よりはやや細分されているが学部相当とみられる(学類の下に、学科にほぼ相当する「コース」が置かれる)。早稲田大学では、分野別の10の「学術院」があり、各学術院の下に、学部、大学院研究科、研究所が存在する。たとえば理工学学術院には3つの学部と6つの研究科が属している。金沢大学の学域と早稲田大学の学術院は性格を異にするが、いずれも代表機関と第2階層下部組織の間に位置する「第1.5階層」ともいべき組織である。

また、九州大学や東京農工大学のように、大学院を教育組織(学生が属する)と研究組織(教員が属する)に分ける大学もある。東北大学、富山大学、広島大学、熊本大学では、一部の研究科が「教育部」と「研究部」に分かれている。

(3) 学部の名称の多様化

図表10は、32大学の学部の名称について、1998年度から2014年度の間の変化を示したものである。2014年度のデータは機関名辞書(Version 2015.1)によるが、1998年度のデータは文献[4]による。ここでは、末尾の「学部」、「学群」等の部分を除き、専攻を示す名称(理、医、工、農など)だけで集計した。この期間に学部の名称が大きく多様化したことが判る。これには、以前存在しなかった新領域の学部の出現と、学部名称の変更(農学部や文学部が別の名称に変わった場合が多い)の両方の影響がある。

図表 10 32大学の学部の名称の分布—1998年度から2014年度の間の変化



(4) 複数の大学による連合大学院研究科の出現

複数の大学により構成される連合大学院研究科(United Graduate School)が各処に設立されている。連合農学研究科が最も多いが、教育学、獣医学、小児発達学等の連合研究科もある。

3-3-2 全学的組織の拡大

(1) いろいろなタイプの全学組織

大学の研究活動の活性化や効率化、可視性の向上、外部との連携、教員や学生の支援のため、全学的組織が大幅に増えている。これらを系統的に分類することは難しいが、次のようなタイプのものがある。

- a) 学部・研究科を横断した研究組織
- b) 全国共同利用型研究施設、共同利用・共同研究拠点
- c) 時限的研究制度・資金によるプロジェクト型組織(世界トップレベル研究拠点プログラム(WPI)等)
- d) 産学協同、地域連携のための組織
- e) 教員や学生の教育・研究・学内生活のための支援組織(保健衛生、環境保全、留学生対応、男女共同参画等)

(2) 全学組織の階層構造

上記の組織の階層付けの仕方は大学によって様々である。大学直属の独立組織、共通性のあるセンターをまとめた機構や本部組織、事務局組織の一環などがある。

多くの大学に共通に存在する全学組織である、情報処理部門、アイトープ管理部門、産学連携部門について、24 の国立大学における組織の名称、上部組織(存在する場合)、組織の種別(大学での位置づけ)を図表 11 に示す。これらの組織の目的や任務は大学によってそれほど大きく異なるとは思えないが、組織的な位置づけや構成は非常に多様である。

3-3-3 頻繁な組織変更

新しい状況への対応のために、組織の新設、改組が頻繁に行われる。時限的な組織が多いこともあるが、学部や研究科の再編、その下の学科や専攻の組織替えも始終行われる。このため、複雑な組織の把握が一層困難になっている。

2012 年 7 月時点で調査した国立 12 大学の下部組織のうち、機関名辞書の最新版(Version 2015.1)でその名称が存続しているものの割合を図表 12 に示す。全体としては、3 年足らずの間に約 15%の下部組織が何らかの形(統廃合、改組、名称の変更等)で変遷していることになる。

図表 11 国立 24 大学における 3 つの全学組織の比較

大学	情報処理部門			アイソープ管理部門			産学連携部門		
	下部組織名称	上位組織	下部組織種別	下部組織名称	上位組織	下部組織種別	下部組織名称	上位組織	下部組織種別
北海道大学	情報基盤センター		全国共同利用施設	アイソープ総合センター		学内共同教育研究施設等	産学連携本部		運営組織・その他
東北大学	教育情報基盤センター		学内共同教育研究施設	サイクロロン・ラジオアイソープセンター		学内共同教育研究施設	産学連携推進本部		特定事業組織
筑波大学	学術情報メディアセンター		学内共同教育研究施設等	アイソープ環境動態研究センター		学内共同教育研究施設等	国際産学連携本部		学内共同教育研究施設等
群馬大学	情報基盤部門	総合情報メディアセンター (他に図書館部門)		RI研究棟	医学部	学部	産学連携・知的財産戦略室	研究・産学連携戦略推進機構 (他に研究戦略室、機器分析センター)	
千葉大学	統合情報センター			アイソープ実験施設			産学連携研究推進ステーション	学術研究推進機構 (他に研究推進部門等)	
東京大学	情報基盤センター		全学センター	アイソープ総合センター		全学センター	産学連携本部		本部業務を行う室等
東京医科歯科大学	情報基盤部門	図書館情報メディア機構 (他に図書館、メディア教育推進部門)		アイソープ部門	歯科学研究支援センター (他に機器分析部門等)		研究・産学連携推進機構		
東京工業大学	情報基盤支援センター	技術部		放射線総合センター		学内共同教育研究施設	産学連携推進本部		企画立案組織
東京農工大学	総合情報メディアセンター			放射線研究室			先端産学連携研究推進センター		
新潟大学	情報基盤センター	学術情報基盤機構 (他に図書館等)	機構	アイソープ総合センター	研究推進機構 (他に機器分析センター等)	機構	産学地域連携推進機構		機構
富山大学	総合情報基盤センター		学内共同教育研究施設	アイソープ実験施設	生命科学先端研究センター (他に動物実験施設等)	学内共同教育研究施設	産学連携部門	地域連携推進機構 (他に生涯学習部門等)	機構
金沢大学	総合メディア基盤センター		学内共同教育研究施設	アイソープ総合研究施設; アイソープ理工系研究施設	学際科学実験センター (他に機器分析研究施設、実験動物研究施設等)	学内共同教育研究施設	産学官連携・知財推進グループ	先端科学・イノベーション推進機構 (他に研究推進グループ)	その他附属施設
信州大学	総合情報センター		附属施設	放射線同位元素利用部門	学術研究推進機構ヒト環境科学研究支援センター (他に機器分析部門、動物実験部門等)	機構	産学官・社会連携推進機構		機構
岐阜大学	総合情報メディアセンター		共同教育研究支援施設	RI管理室	生命科学総合研究支援センター (他に機器分析分野、動物実験分野等)	共同教育研究支援施設	産学官連携推進本部	研究推進・社会連携機構 (他に研究推進部門、地域連携部門)	
名古屋大学	情報基盤センター	情報連携統括本部	運営支援組織; 共同利用・共同研究拠点	アイソープ総合センター		学内共同教育研究施設等	学術研究・産学官連携推進本部		運営支援組織
京都大学	学術情報メディアセンター		教育研究施設等	放射線管理部門	環境安全保健機構 (他に環境管理部門、安全管理部門等)	教育院等	産学官連携本部		教育院等
大阪大学	サイバーメディアセンター		全国共同利用施設	ラジオアイソープ総合センター		学内共同教育研究施設	産学連携本部		
神戸大学	情報基盤センター		学内共同教育研究施設等	アイソープ部門	学術研究推進機構研究基盤センター (他に機器分析部門等)	機構	連携創造本部		学内共同教育研究施設等
岡山大学	情報統括センター		全学センター	光・放射線情報解析部門	自然生命科学研究支援センター (他に動物資源部門等)	全学センター	産学官連携本部	研究推進産学官連携機構 (他に研究推進本部、社会連携本部等)	
広島大学	情報メディア教育研究センター		学内共同教育研究施設等	アイソープ総合部門	自然科学研究支援開発センター (他に低温・機器分析部門等)	学内共同教育研究施設等	産学・地域連携センター		学内共同教育研究施設等
徳島大学	情報センター		共同教育研究施設等	アイソープ総合センター		共同教育研究施設等	産学官連携本部; 四国産学官連携イノベーション共同推進機構		共同教育研究施設等
九州大学	情報基盤研究開発センター		共同利用・共同研究拠点	アイソープ総合センター		学内共同教育研究施設	産学官連携本部; 産学連携センター		推進室; 学内共同教育研究施設
長崎大学	ICT基盤センター			アイソープ実験施設	先端生命科学支援センター (他に動物実験施設等)		産学官連携戦略本部		
熊本大学	総合情報統括センター		学内共同教育研究施設	アイソープ総合施設	生命資源研究・支援センター (他に遺伝子実験施設等)	学内共同教育研究施設	ベンチャービジネスラボラトリー	イノベーション推進機構	

図表 12 国立 12 大学における下部組織の変遷－2012 年 7 月に存在する下部組織の 2015 年 2 月における存続状況

大学	2012年度の 下部組織数	うち2015年 2月に現存	存続率 (%)
北海道大学	69	68	98.6
東北大学	70	59	84.3
筑波大学	93	62	66.7
千葉大学	49	42	85.7
東京大学	53	52	98.1
東京工業大学	91	74	81.3
名古屋大学	60	56	93.3
京都大学	74	67	90.5
大阪大学	72	57	79.2
神戸大学	46	42	91.3
広島大学	63	53	84.1
九州大学	125	104	83.2
総計	865	736	85.1

3-4 大学下部組織情報の収集の難しさ

機関名辞書には、32 大学の下部組織(原則として第 2 階層まで)を網羅的に収録することとしている。また、現存しない下部組織についてもできるだけ収録し、現存下部組織との継承関係を示すこととしている。そのため、定期的にこれらの下部組織を特定し、その日本語と英語の名称を知る必要がある。

現在のところ、これらの情報を得るための最も確実な情報源は、各大学の Web サイトにおける組織図や組織案内のページである。しかし、これらの Web サイトには、次のような問題点があり、3-3 で述べた複雑な組織構造を理解するのに困難があるのが実情である。学部及び大学院研究科の構成についての情報は比較的容易に得られるが、それ以外の全学的組織についての理解が特に難しい。

(1) Web サイト情報における下部組織の欠落

Web サイトにおける組織図や組織案内のページに載っていない下部組織が存在する。時限的な組織や学生・教員支援組織にこのような例が多い。この場合は、その下部組織のホームページを探すが、それもないことがある。

(2) 組織の階層関係の不明確さ

Web サイトに示されている組織図や組織案内ページでは、第 2 階層下部組織とその下の下部組織が並列に表示されているなど、階層関係が必ずしも明らかでないことがある。このような場合、機関名辞書への収録対象である第 2 階層下部組織の識別が困難である。

(3) 下部組織英語名を知ることの難しさ

Web サイトの組織図に下部組織英語名が併記されている大学はほとんどない(各大学 URL から情報を

取得した 2015 年 2 月時点における 32 大学中唯一の事例は富山大学)。これ以外のページに日英併記の表がある大学が少数ある(但しそれを見つけ出すことは容易ではなく、また、データが更新されていないことがある)。それ以外のほとんどの大学では、次のどちらかの方法で英語名を探すことになる。

- ① English サイトの組織表を見る。
- ② 下部組織のホームページを見る。

①では、英語の組織表があっても、日本語のそれと対応していなかったり(抜けている下部組織がある、順序が対応していないなど)、バージョンが古かったりすることがある。②では、英語名記載のないホームページがある。このような状況では、同じ大学の中に似た名称の下部組織があると、日本語名と英語名の対応づけに困難が生ずる。

(4) 現存しない下部組織の探索

論文データベースには、廃止、改組、名称変更等により現存しない下部組織のデータも含まれるので、その同定も必要であるが、大学の Web サイトには非現存組織の情報はほとんど存在しないので、これはきわめて困難である。機関の沿革のページから変遷関係を知ることができる場合があるが、その場合でも、非現存下部組織の英語名はまず判らない。従って、少なくとも年に一度、定期的にその時点での下部組織調査を行い、機関名辞書を更新していく必要がある。

4 NISTEP 大学・公的機関名辞書を用いた論文データベースの機関同定

NISTEP では、機関名辞書の適用例として、この辞書と論文データベースをリンクさせることによって、研究活動を行っている日本の機関の研究活動データの取得と分析を行っている。その際重要なのは、論文データベース中の著者所属機関データと機関名辞書中の機関(代表機関または下部組織)との対応付け(機関同定)である。この章では、Web of Science Core Collection (WoSCC)データベースを用いた機関同定について、同定を難しくする問題点を論じた後、同定処理方法の概要及び問題点に対処するための工夫について述べる。

4-1 論文書誌情報(著者所属機関表記)に基づく機関同定の概要とその難しさ

我が国の機関を含む論文(著者の少なくとも一人が日本の機関に所属する論文)を論文データベースから抽出し、その機関が機関名辞書に収録されるどの機関に該当するか(あるいは機関名辞書に含まれない機関に該当するか)を同定するには、以下の手順による。ここに示したのは論理的手順で、実際の同定アルゴリズムでは②と③は並列的に行われる。

- ① 論文データベースの著者所属機関フィールドから、日本の機関を抽出する。通常、所属機関フィールドには機関が所属する国名が含まれているので、これを利用して抽出する。
- ② 所属機関の表記と機関名辞書との文字列マッチングにより、その表記に該当する機関(代表機関)を同定する。
- ③ ②でマッチした代表機関の下部組織が機関名辞書に含まれている場合には、更に、それらの下部組織名と所属機関表記の残り部分とを照合し、マッチする下部組織があればそれに同定する。
- ④ マッチする代表機関がない場合は、所属機関表記中の特徴語(“Corp”等の機関接尾辞)を用いて、セクター分類もしくは病院フラグの付与を行う。これらの付与もなされない所属機関表記に対しては同定不能のフラグを与える。

しかし、この同定処理はそれほど単純ではない。機関名辞書に存在する機関であるのに上記の④のように同定に失敗(同定漏れ)する場合、本来の機関とは別の機関に誤同定される場合等がある。事後のサンプリング調査でこのような誤りが判明した場合は、機関名辞書への新機関の登録や既存機関への別名の追加、同定アルゴリズムの見直し等の検討を行う。

同定漏れや誤同定が起こるのは、主に次のような理由による。これについては参考文献[2]の 4-4～4-5 も参照されたい。なお、以下の例示は WoSCC のデータによるため、そこで用いられている短縮語表記と前置詞、接続詞等の省略がなされている。

4-1-1 機関名の表記ゆれ

論文データベースの所属機関には、同じ機関の名称が様々の形で表記されている(これは、もとの論文における著者所属機関の表記が多様であることの反映である)。そのパターンは次のように大別され

る。

(1) 正式の名称とは単語や語順が異なる表記

東京大学(正式名(the) University of Tokyo)を Univ Tokyo ではなく Tokyo Univ と表記、東京農工大学(正式名 Tokyo University of Agriculture and Technology)を Tokyo Univ Agr & Technol でなく Tokyo Noko Univ と表記など。

(2) 語の間違いや脱落

東京農工大学(Tokyo University of Agriculture and Technology)が Tokyo Univ Agr & Engrn あるいは Tokyo Univ Agr と誤記されていると、東京農業大学(Tokyo University of Agriculture)に誤同定されてしまう。このように、ちょっとした語のミスで同定不能や誤同定が起こる。

(3) 単語の略記

WoSCC の所属機関データでは、University→Univ、Institute→Inst、Medicine (Medical) →Med、Science (Scientific) →Sci などの略記が用いられるが、それほど一般的でない語も略記されているため同定漏れになることがある。

(4) 機関の略称

産業医科大学(University of Occupational and Environmental Health)に対する UOEH、総合研究大学院大学(The Graduate University for Advanced Studies)に対する GUAS、SOKENDAI など。

(5) 冠詞、前置詞、接続詞、記号(ハイフン等)の省略や書き換え

論文中ではこの種の表記ゆれが多数あるが、WoSCC ではこれらの品詞はすべて省略される(and は&に変換)ので、これに関する問題は少ない。しかし、WoSCC ではハイフンを省略するとき、その間を詰める場合(Electro-Commun→Electrocommun 等)と開ける場合(Bio-Appl→Bio Appl)があり、注意が必要である。

(6) 機関の旧名の表記

たとえば、東京薬科大学(Tokyo University of Pharmacy and Life Science)、東京理科大学(Tokyo University of Science)、自治医科大学(Jichi Medical University)に対して、旧名の Tokyo Coll Pharm、Sci Univ Tokyo、Jichi Med Sch もよく使われている。

(7) スペルの違い

単数形と複数形のゆれ、へボン式と訓令式のゆれ(Kyushu と Kyusyu 等)、単語間のハイフン挿入(Kwansei Gakuin と Kwansei-Gakuin 等)、単純なスペルミスなどがある。また、府立大学や県立大学の府県を示す単語の表記には、Prefectural、Prefectoral、Prefectual、Prefecture と様々なゆれがある。

4-1-2 代表機関名と下部組織名のサブフィールド配置の不規則性

WoSCC では、所属機関フィールド(機関名の他、所在地、郵便番号、国の情報を含む)から代表機関と下部組織の部分を取り出し、それぞれを ORG、SUBORG というサブフィールドに収録している。これは同定を行う上で便利であるが、必ずしもこの通りになっていないことがある。

そのひとつとして、ORG サブフィールドに代表機関と下部組織あるいはキャンパス名が結合した表記が見られる。東京大学の場合、Univ Tokyo Hosp、Univ Tokyo Grad Sch Med、あるいは Univ Tokyo Hongo、Univ Tokyo Komaba 等がある。しかしこれらは、下部組織同定の手続きがやや複雑化するものの大きな問題ではない。より深刻なのは、ORG サブフィールドと SUBORG サブフィールドに代表機関名が分離してしまう場合である。たとえば、獨協医科大学(Dokkyo Medical University)は Dokkyo University School of Medicine と表記されることもよくあるが、その Dokkyo University が ORG サブフィールドに、School of Medicine が SUBORG サブフィールドに別れていると、ORG サブフィールドの表記から獨協大学(Dokkyo University)に同定されてしまう。獨協大学に医学部はないので、SUBORG サブフィールドを見れば誤同定と判るが、これをアルゴリズムで認識することは容易ではない。

この他に、ORG サブフィールドに下部組織名が、SUBORG サブフィールドに代表機関名が入っている場合もある。

4-1-3 一つの所属機関フィールドに複数の機関が存在

論文の著者が複数の機関に所属していることがある。このような場合、論文の著者所属機関には複数の機関が併記されるのが通例なので、データベースにおいても別々の所属機関フィールドに収録される。しかし、著者が複数の機関(または組織)を合体して論文に記述していると、データベースでもそれらが分離されず、一つの機関データに2機関が存在してしまう場合がある。次の2つの例の a)では東京大学物性研究所と科学技術振興機構(JST)の事業である CREST が、b)では国立遺伝学研究所と総合研究大学院大学が合体表記されている。

- a) Inst Solid State Phys, Univ Tokyo, JST-CREST, Japan
- b) Div Mammalian Dev, Natl Inst Genetics, Dept Genetics, SOKENDAI

この著者はその両方の機関に所属していると考えられるので、両方の機関を同定できるアルゴリズムが必要である。WoSCC の場合、両方の機関が ORG サブフィールドに入っている場合と、一方が ORG サブフィールドに、他方が SUBORG サブフィールドに入っている場合がある。

4-1-4 同一の英語機関名を持つ機関が存在

無関係の機関が同一の英語名を持つ例は、大学や公的機関ではほとんどないが、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それ程珍しくない。このような場合、機関データの英語名が新旧どちらの機関に対応するか判断しづらい。代表的例では、東京都立大学と

首都大学東京はどちらも Tokyo Metropolitan University である。また、短期大学から4年制大学への移行、国立研究機関から独立行政法人への改組の際にも、旧英語名がそのまま保存されていることがある。後者の例として、宇宙科学研究所は独立した国立研究所から独立行政法人の宇宙航空研究開発機構の下部組織に移行したが、Institute of Space and Astronautical Science の英語名をそのまま用いている。

4-1-5 英語名が非常に似ている機関が存在

機関名が機関表記によく使われる単語のみから成る場合、類似した名前の機関が存在して誤同定が起りやすい。分子科学研究所(Institute for Molecular Science)は、大学共同利用機関自然科学研究機構に属する研究所であるが、論文では機構の名称は略されることが多い。ところが、Inst Mol Sciあるいはこれとごく類似した名称の附属研究所を持つ機関は多数存在する(たとえば愛知医科大学の分子医科学研究所(Institute of Molecular Medical Sciences))。そのほかの例として、公益財団法人がん研究会のがん研究所(単に研究所名の Canc Inst で表示されることが多い)、旧・国立公衆衛生院(The Institute of Public Health)(現・国立保健医療科学院)などが挙げられる。

図表 13 英語名が似通っているため同定に注意を要する大学の組み合わせ

組合せ番号	大学A	大学B	大学C	大学Aの英語名	大学Bの英語名	大学Cの英語名
1	静岡大学	静岡県立大学		Shizuoka University	University of Shizuoka	
2	東京理科大学	山口東京理科大学	諏訪東京理科大学	Tokyo University of Science	Tokyo University of Science, Yamaguchi	Tokyo University of Science, Suwa
3	獨協医科大学	獨協大学	姫路獨協大学	Dokkyo Medical University	Dokkyo University	Himeji Dokkyo University
4	浜松医科大学	浜松大学 (現・常葉大学)		Hamamatsu University School of Medicine	Hamamatsu University (Tokoha University)	
5	長崎大学	長崎県立大学		Nagasaki University	University of Nagasaki	
6	高知大学	高知県立大学		Kochi University	University of Kochi	
7	兵庫県立大学	兵庫大学		University of Hyogo	Hyogo University	
8	金沢大学	金沢医科大学		Kanazawa University	Kanazawa Medical University	
9	埼玉大学	埼玉医科大学		Saitama University	Saitama Medical University	
10	岩手大学	岩手医科大学		Iwate University	Iwate Medical University	
11	関西大学	関西医科大学		Kansai University	Kansai Medical University	
12	和歌山大学	和歌山県立医科大学		Wakayama University	Wakayama Medical University	
13	福島大学	福島県立医科大学		Fukushima University	Fukushima Medical University	
14	愛知大学	愛知医科大学		Aichi University	Aichi Medical University	
15	奈良県立医科大学	奈良大学		Nara Medical University	Nara University	
16	札幌大学	札幌医科大学		Sapporo University	Sapporo Medical University	
17	島根大学	島根県立大学	島根県立大学短期大学部	Shimane University	The University of Shimane	The University of Shimane Junior College

大学名が似通っているため同定が難航する例を図表 13 に示す。データベース中の表記が誤っていたり曖昧であったりすると、違う大学に同定されてしまう(上記[2]で述べた獨協医科大学の例はその一例)。表記の間違ひは、論文自身の表記が間違っている場合と、データベース収録の際のミスの方が

あるようである。静岡大学(Shizuoka University)と静岡県立大学(University of Shizuoka)の場合、どちらも静岡市駿河区にあり(静岡大学は浜松市にもキャンパスがあるが)、郵便番号の最初の3桁は共通で、やや似通った学部が存在するので、大学名の表記が間違っていると識別は極めて困難である。

4-2 大学の下部組織同定の問題点

3-1 に示した 32 大学については、その下部組織を機関名辞書に網羅的に収録し、下部組織レベルの同定を行っている。この場合も、4-1-1~4-1-5 に示した理由による同定の難しさが存在する。

表記のゆれは機関名よりも更に著しい。たとえば、東京工業大学の総合理工学研究科の正式英語名は Interdisciplinary Graduate School of Science and Engineering であるが、WoSCC には以下のように種々の表記がある。

Interdisc Grad Sch Sci & Engr (正しい表記)

Interdisc Grad Sch Sci Engr

Interdisc Grad Sch Engr

Interdisc Grad Sch Engr & Sci

Interdisc Grad Sch Sci & Tech

Grad Sch Interdisc Sci & Engr

また、下部組織名が一般にはあまり知られていない略称で表記されていることがある。このほかに、下部組織同定に独自な問題点として、次のようなものがある。

4-2-1 多くの大学に共通の下部組織名の存在

これは当然のことなので問題点と言うには当たらないが、Faculty of Medicine、Graduate School of Science、University Hospital 等は多くの大学に共通して存在する下部組織名である。従って、下部組織レベルの同定は、その上位の大学レベルでの同定を行った後、その下位に存在する下部組織と照合することが必要である。

4-2-2 第2階層より下の組織名での所属表記

論文中の所属表記においては、大学直下の第2階層下部組織を省略して第3階層下部組織が記載されることがしばしばある。つまり、学部名や研究科名(第2階層下部組織)を記載せず、直接学科名や専攻名(第3階層下部組織)が記載されているようなケースである。確かに、Graduate School of Science, X University よりも Department of Biochemistry, X University の方が著者の所属や研究分野を具体的に示すので、このような所属表記がなされるのであろうが、これも下部組織同定を難しくする一因になる。大きな大学では、Department of Biochemistry は理学部、医学部、薬学部、農学部のいずれにも存在し得る。

学部や学科が安定して存在しておればともかく、大学の組織は始終改組が行われるので、問題は深刻である。

4-2-3 大学教員の複数下部組織所属

通常、大学には学部(Faculty または School)と大学院研究科(Graduate School)があり、多くの教員はこの両方に所属している。最近ではこの他に、教員が本籍を置く組織を設ける大学が増え、これは研究院、学系などと呼ばれる(英語では Research Faculty、Institute など)。更に、その他のプロジェクト型組織にも所属することがある。これらに重複して所属する教員が、論文発表の際どの所属組織を記載するかは、人により、あるいは同じ人でも時期によって様々と思われる。論文発表の際に記載する組織が一定していないことは、同定よりも集計、分析を行う場合の問題点である。

4-3 機関同定アルゴリズムの骨子

4-1 では、WoSCC を例として、所属機関データを機関名辞書の代表機関または下部組織に同定する手順の概要を述べた後、同定処理を困難にする所属機関データ表記上の問題点を列挙した。また、4-2 では、大学下部組織のデータ表記に特有の問題点を示した。

NISTEP では、これらの問題点を踏まえて、代表機関・下部組織同定を行うためのプログラム、NISTEP 論文機関名同定プログラム(Web of Science バージョン)を開発してきた。この節ではそのアルゴリズムの概要を示す。4-1 のはじめに機関同定処理手順の概要を示したが、アルゴリズムとしては次の流れになる。

4-3-1 前処理－データの正規化

データベース中の所属機関データ、機関名辞書内の機関名(英語)データを以下のように規格化した後(WoSCC のデータではこれらの規格化の多くが既になされている)、スペースで区切った単語列で表現する。

- ① アポストロフィ“ ’ ”の除去
- ② “/”と“&”を“and”に変換
- ③ 指定した前置詞、冠詞、接続詞の削除
- ④ ハイフン“-”の除去: 除去した部分を詰めたデータと空白としたデータの両方を生成(WoSCC にこの両方の処理があるため)
- ⑤ カンマ“ , ”の処理: カンマを 1 つの単語と見た場合と削除した場合の両方の単語列を生成⁹

⁹ 所属機関データにはカンマが省略されていることが多いため、以下の 4-3-2 で述べる最長マッチングでは、カンマを無視してマッチング処理を行っていたが、これで不都合が生じる場合がある。たとえば、“Tokyo Inst Technol, Dept Phys”という文字列があった時に、“Technol, Dept”の部分が Technical Department(東京工業大学技術部)にマッチしてしまう。このような誤同定を防止するため、機関

- ⑥ 略記辞書の適用:University→Univ、Science(s)→Sci 等、400 弱の単語に対する略記が存在
- ⑦ 4 文字以上の単語の末尾が ogical, ogy, ogies, ical, ic, ics, ion, ions のときその部分を削除

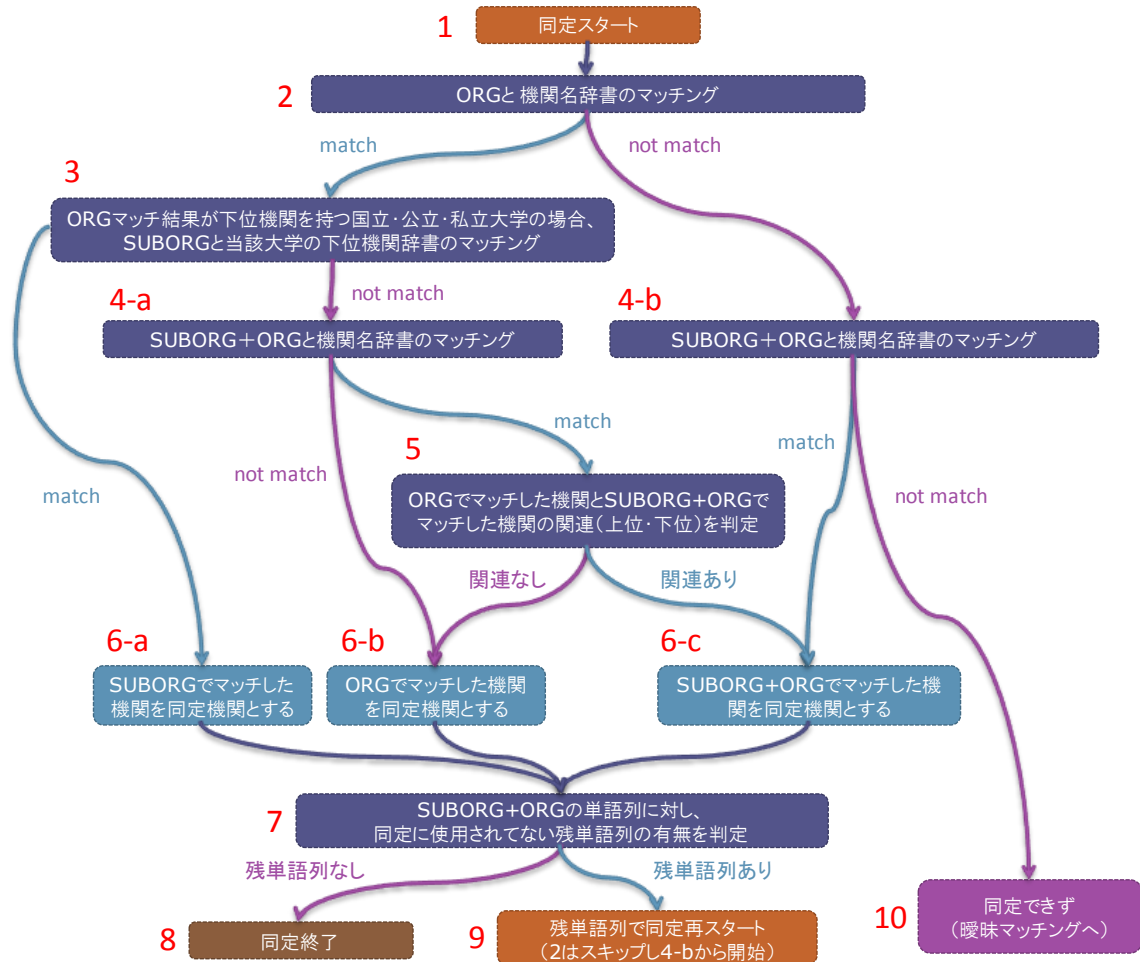
4-3-2 最長マッチング処理

同定処理の中核部分であり、現在のところ、WoSCC の 1996-2013 年(出版年)データを対象とした同定処理で、処理対象機関データの 92%、機関同定されるデータの 98%がこの処理により同定される。4-3-1 の前処理を行った所属機関データ表記単語列を機関名辞書内表記単語列と比較し、後者を完全に内包する単語列が前者の中に見つかれば、そのうち最も長い単語列のデータに相当する機関名辞書の機関に同定する(同長の単語列が複数見つかった場合は、場合によりそのいずれかまたは両方を同定する)。

WoSCC では、所属機関データのうち代表機関と下部組織の部分をそれぞれ ORG、SUBORG というサブフィールドに収録している(4-1-2 参照)が、上記のマッチング処理はまず ORG サブフィールドに対して行い、その後、ORG で同定された機関が大学である場合は SUBORG に対して、それ以外の場合は SUBORG と ORG を合体した単語列(SUBORG+ORG)に対して行う。この手順をとる主な意図は、まず代表機関を、次に下部組織を同定して、同定された下部組織が代表機関に属するときは下部組織の同定を優先することであるが、別の意味もある。一つは、SUBORG+ORG 単語列とのマッチングにより、4-1-2 で述べた状況(ORG と SUBORG へのデータ配置が原則から外れている場合)にもある程度対処できることである。もう一つは、4-1-3 で述べた、一つの所属機関フィールドに複数機関が存在する場合を発見するためである。最長マッチングのフローを図表 14 に示すが、この図のステップ 9 により、SUBORG+ORG 単語列を逐次的にサーチする。

名辞書のデータを読み込む際に、カンマのある・なしの両方の探索パターンを作成するようにした。上記の例では、“Tokyo”, “Institute”, “Technology”, “,”, “Department”, “Physics”と、“Tokyo”, “Institute”, “Technology”, “Department”, “Physics”の 2 つのパターンが作成される。最長マッチングにより前者から“Tokyo Inst Technol”が優先同定される。

図表 14 最長マッチングの流れ



4-3-3 曖昧マッチング処理と郵便番号マッチング処理

最長マッチングでは所属機関データの同定ができなかった場合、その文字列に対して、曖昧マッチング(N-gram¹⁰とレーベンシュタイン距離¹¹を使用したマッチング)を適用する。現在のアルゴリズムでは、所属機関データ中に機関名辞書データの文字列と1文字違いの文字列が発見された場合に曖昧マッチが成立する。それでも同定がなされない場合、所属機関のアドレスデータに郵便番号が含まれていれば、機関名辞書中の郵便番号データとのマッチングによる同定を行う。

2014年度に行った同定処理の結果をサンプリング調査することにより、これらのマッチング処理が有効

¹⁰ 機関名辞書中のデータをN字(たとえばN=3)ごとに区切った文字列の索引を作り、所属機関データに含まれる文字列を高速検索する方法。

¹¹ N-gramでマッチした候補データについて、挿入、欠失、置換を考慮した比較によりマッチングを行う方法。

に機能する場合もあるが、一方においてかなりの誤同定が発生する(曖昧マッチング、郵便番号マッチングでそれぞれ、同定されたデータの約 30%、約 40%)ことが判った。曖昧マッチングでは、辞書未登録の簡単な名称の会社が1字のみ異なる辞書中の会社に同定されたり、機関接尾辞(“Co., Ltd.”、“Corp”、“Inc”等)の誤表記のため別の会社に同定されたりするケースが多かった。郵便番号マッチングでは、大きな大学や研究機関の敷地内にベンチャー的な機関が所在する場合や、研究団地のビル内に多くの企業が同居している場合に誤同定が起りやすい。

このようなことが判ったので、今後は、曖昧マッチング処理と郵便番号マッチングをそれぞれ独立に適用することはやめ、両者の結果が一致するときに同定が成立するように条件を変更する考えである。また、調査の結果、辞書の増強やアルゴリズムの軽微な修正により、最長マッチングで処理できるケースがかなりあることが判明した。そもそも、大部分のデータが最長マッチングで同定され、曖昧マッチングと郵便番号マッチングで同定される割合は合わせて 1%程度なので、以上の修正の結果同定漏れとなるデータはそれほど多くはない。

4-3-4 機関同定できなかったデータの処理

以上の処理で機関同定ができなかったデータに対しては、次の処理を行う。

- ① 機関は同定できないがセクターが判明するデータにはセクター分類を与える。たとえば、データに“Co., Ltd.”が含まれていれば会社セクターに分類する。
- ② 機関は同定できないが“Hospital”等の文字列が含まれているデータに対して病院フラグ(4-4を参照)を付与する。
- ③ それ以外のデータは、日本の機関であるという情報のみが残る。

2014年度に行った WoSCC データの処理では、機関同定ができなかったのは全データの 6.2%で、そのうち、セクター分類ができたもの、病院フラグが付与されたもの、全くの同定漏れになったものが、それぞれ 1.8%、2.8%、1.7%であった。

機関同定漏れとなったデータのうち出現頻度がある程度高いものについては、正しい機関を調査し、補助セットへのデータ追加やマッチングアルゴリズムの修正等を適宜行っている。しかし、このような措置が別のエラーを招くこともあるので、同定漏れよりも誤同定を防止することを優先して、慎重に進めている。

4-4 機関名辞書の補助セット

2-3 で述べたように、機関名辞書はコアセットと補助セットから構成される。補助セットには、コアセットへの収録基準を満たさないが同定や分析の作業上必要性が高い機関や情報を主に収録する。4-3 で述べた同定処理の際には、コアセットと補助セットの間に優先順位を付けることなく、両者を合体して用いる。

補助セットに収録される機関と情報は、具体的には次の通りである。

(a) コアセットに収録される機関・下部組織の正式名称以外の名称(日本語及び英語の別名、略称等)

2-3 に示したコアセットに収録される代表機関と下部組織について、正式名称以外の名称(日本語及び英語の別名、略称等)を収録する。

(b) 以下の機関・組織に関するすべての情報(機関ID、日本語及び英語の正式名称とそれ以外の名称、セクター、上位機関とのリンク、変遷情報、所在地情報)

① 会社(代表機関、下部組織とも)

② 以下の機関の第2階層下部組織

- ・ 大学下部組織のうちコアセットに含まれる以外のもの(北里大学北里生命科学研究所等)
- ・ 国の機関及び特殊法人・独立行政法人等に所属する下部組織のうちコアセットに含まれる以外のもの(農林水産省動物検疫所、独立行政法人産業技術総合研究所関西センター等)
- ・ 地方公共団体の機関、非営利法人、その他の機関の下部組織

③ 第3階層下部組織

機関IDの形式は、コアセットと補助セットで共通とする。補助セット中の下部組織については、コアセット中に存在する該当の代表機関(第3階層組織の場合はその上位の第2階層組織)へのリンク情報を与える。

(c) 病院フラグ

当該の代表機関または下部組織が病院である場合、このフラグを付与する。その機関・組織がコアセット、補助セットのいずれにあるかを問わない。病院であることを定める基準が確定していないので、当面この情報は補助セットに収録する。

補助セットに含まれる情報のうち、機関同定に直接関係が深いのは英語の別名・略称である。これらは、同定漏れや誤同定の原因(4-1、4-2参照)を防止するために収録されている場合が多いので、収録に明確な基準があるわけではない。従って、内部作業用に用い、公開の機関名辞書には含めていないが、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」には含まれている。

4-5 同定漏れや誤同定を防ぐための具体例

上述のように、補助セットの充実やマッチングアルゴリズムの改善により、同定漏れや誤同定をできるだけ減らし、同定精度を上げる努力をしているが、ここでは、いくつかの個別の例を示す。

4-5-1 データに複数機関が併記されている場合

論文の著者が複数の所属機関(または組織)を記述している場合、WoSCC等の論文データベースでは、通常それらの機関データごとに別々のレコードを作成する。しかし、4-1-3で述べたように、論文に著

者が複数の機関を合体して記述していると、データベースでもそれらが分離されず、一つの機関データに複数の機関が存在することがある。

このような場合に複数の機関を漏らさず同定できるように、同定アルゴリズムでは4-3-2で述べたような処理を行っている。すなわち、SUBORG+ORG 単語列で一つの機関がマッチした後、単語列の残りの部分が別の機関にマッチしないかサーチし、全単語列をサーチするまでこの手続きを繰り返す。

4-5-2 英語名が同名の異なる機関

機関名辞書には、英語名が同名の異なる機関が存在する。これには次の2つの場合がある。

- a) 2つの機関の間に継承関係がある(この事例は4-1-4を参照)
- b) 無関係の機関の英語名がたまたま一致

同定の過程でこのような機関にマッチングした場合、まずそれらの同名機関の間に継承関係がある(どちらかの機関が統合や日本語名変更により他方に継承される)か否かを機関名辞書の変遷情報から確認し、それにより以下の処理を行う。

(1) 継承関係がある場合

機関名辞書の移行年と論文の発行年を比較し、発行年が移行年以前なら旧機関に、そうでなければ新機関に同定する(移行年データがない場合は新機関に同定する)。但し、論文の公表は投稿よりかなり遅れることもあるので、この措置は完全ではない。

また、この措置が必ずしも正しくない場合がある。たとえば、機関Aが機関Bに移行し、それに伴いAの下部組織aはBの下部組織bとなったが、aとbの英語名は同じであるとする。そして、A、B、bは機関名辞書に登録されているがaは登録されていなかった場合、「A、a」という機関データ表記は旧機関Aと新機関B-bの両方にマッチし、上記の原則ではB-bに同定されるが、むしろAへの同定が妥当と考えられる。このようにいろいろな場合が起こりえるので、継承機関と上下関係のパターンに応じてきめ細かく同定機関を選別できるようにした。

(2) 継承関係がない場合

無関係の機関の英語名が一致する例は会社や非営利団体に多い(漢字は異なるがローマ字綴りでは同一等)。これに対しては次のように対処している。

- ① ほとんど論文に出現しない機関や所在がインターネットで確認できない機関を機関名辞書から削除。
- ② 郵便番号マッチングの適用。
- ③ 郵便番号マッチングが適用できない場合は機関同定不能としてセクター分類のみ行う(4-3-4を参照)。

4-5-3 英語名が似ている機関

これには次の場合があり、いずれも誤同定を起こしやすい。

- a) 英語名がありふれた単語からなる機関
- b) よく似た英語名の大学

それぞれの具体例を 4-1-5 に示した。これらについては、以下の方法で誤同定を防止に努めている。

(1) 英語名がありふれた単語からなる機関

以下の方法で対処している。

- ① 類似の名称の機関(下部組織を含む)を補助セットに登録し、それらに用いられるいろいろな表記を別名として収録する。
- ② 同定結果をマニュアルで調査し、論文に出現しない機関の英語名を機関名辞書から削除する。

(2) よく似た英語名の大学に対する特別措置

図表 13 に示した大学の組み合わせでは、英語名称が似ているため、データベース中の機関表記でしばしば混同が見られる。これらの誤同定をできるだけ抑えるため、以下の手順による特別措置を執っている。以下、これらの大学を「特別措置大学」という。

① 特別措置大学の郵便番号の調査と辞書の更新

特別措置大学のいずれかに同定された機関データから郵便番号を抽出し、それらが実際にどちらの大学のものであるかを調査した。

② 上記郵便番号情報を用いて特別措置大学用の統計辞書を作成

正しい郵便番号を用いていずれかの大学に帰属された機関データの SUBORG サブフィールドに含まれる単語列から、その大学との共起頻度と共起率(他の特別措置大学を含めた共起頻度のうちその大学との共起の割合)が高いものを取り出し、単語列と大学を関係づける辞書を作成する。たとえば静岡大学と静岡県立大学の場合、“Fac Sci”や“Fac Agr”は前者、“Sch Pharm Sci”や“Sch Food Nutr Sci”は後者に関係づけられる。

③ 機関同定アルゴリズムの改良

最長マッチングで特別措置大学が同定された際、次のプロセスに進む前に、SUBORG サブフィールドの表記と上記統計辞書を比較し、同定結果が統計辞書と矛盾していれば入れ替える処理をアルゴリズムに実装した。

4-6 大学の下部組織同定率向上策

機関名辞書に下部組織情報を網羅的に収録している 32 大学(3-1 参照)については、2012、2013 年度にも下部組織同定を行っていたが、これまでに述べた手順によって下部組織レベルで同定される割合は低く、大学間の下部組織同定率の差も大きかった。そこで、2013 年度の同定結果を 2014 年度において詳細に分析し、下部組織同定率とその精度を上げるための検討を行った。その結果は、これまでに述

べた措置にも反映されているが、ここでは、特に 32 大学の下部組織同定率の向上のために行った改善策について記す。

4-6-1 アドレス情報全体を対象としたマッチング処理

これまで述べてきたように、WoSCC に対する機関同定(最長マッチ・曖昧マッチ)は ORG 及び SUBORG サブフィールドを対象に行っている。しかし、調査の結果、所在地サブフィールド (ADDRESS) にも下部組織の文字列が入っていることがある(特に出版年が 1997 年以前のデータ)ことが判ったので、32 大学については、ORG、SUBORG に加えて、ADDRESS も機関同定に使用することとした。

4-6-2 下部組織統計辞書の開発

下部組織統計辞書の作成手順は次の通りである。

- ① 最長マッチングの結果から、32 大学の下部組織を同定したレコードを抽出する。
- ② 各レコードの SUBORG サブフィールドに含まれる単語列から、大学別に、出現頻度の高いものを抽出する。
- ③ これらの高出現頻度単語列に対して、下部組織別の共起率(大学全体の出現頻度のうちその下部組織との共起の割合)を求める。
- ④ 出現頻度が高く、かつある特定の下部組織に共起が集中している単語列を、その下部組織に関係づける。これが下部組織統計辞書である。

たとえば、東京大学に同定されたデータに“Ctr Human Genome”や“Mol Med Lab”が含まれれば医科学研究所、“Dept Ophthalmol”や“Dept Surg”が含まれれば医学系研究科である確率が高いとして、それらの下部組織に関係づける。

現在の下部組織統計辞書には、次の条件のいずれかを満たす約 5,700 の単語列を収録している。

- a) 出現頻度が 30 以上で、ある下部組織との共起率 0.8 以上
- b) 出現頻度が 10 以上 30 未満で、ある下部組織との共起率 1

こうして作成した下部組織統計辞書を、最長マッチングで 32 大学に同定されたがその下部組織同定はできなかったデータ(機関名辞書に含まれる下部組織名とはマッチしなかったデータ)に適用する。その際は、ORG、SUBORG、または ADDRESS のすべての文字列が対象となる。

4-6-3 補助セットの対象拡張

下部組織統計辞書の対象となった下部組織表記のうち、特定の下部組織を示すことが間違いないと人手で判断できる表記は、機関名辞書の補助セットに加えることとした。補助セットに含まれる機関名称(4-4 参照)は、原則として当該の機関や下部組織自体を示す別名や略称であるが、32 大学の下部組織

についてはそれ以外の表記にも拡張した。たとえば東京農工大学の場合、“Dept Anim Life Sci”は農学研究院(Institute of Agriculture)そのものを示す名称ではないが、同大学に同定された機関データにこの表記があれば、農学研究院に同定して間違いはない。同様に、“Dept Math”の表記は確実に工学研究院(Institute of Engineering)に結びつけられる。同定データからこのような表記を拾い出し、補助セットに加えることにより、下部組織同定率の向上を図った。

4-6-4 大学名と下部組織名の反転によるマッチング

機関名辞書では、大学の下部組織名は、“Graduate School of Medicine, Gifu University”のように、下部組織＋代表機関の順番で記載されている。WoSCCの機関データが、原則通りORGサブフィールドに大学名が、SUBORGサブフィールドに下部組織名が表示されていれば問題ないが、ORGサブフィールドに“Gifu Univ Grad Sch Med”のように記載されている場合がある。このような場合に対処するため、機関名辞書に下部組織情報を持つ大学に限り、機関名辞書通常の語順に加え、大学名と下部組織名を反転させた語順でもマッチングを行うようにした。

図表 15は4-1、4-2に示した同定上の問題点と、4-3～4-6に述べた対策とを対応づけたものである。

図表 15 機関同定上の問題点とその対策のまとめ

問題点	対策	対策を示した章節
機関名の表記ゆれ(4-1-1)	・機関名辞書への別名の増強 ・曖昧マッチング処理	4-4 4-3-3
代表機関名と下部組織名のサブフィールド配置の不規則性(4-1-2)	最長マッチング処理(ORG+SUBORG単語列でのマッチング)	4-3-2
一つの所属機関フィールドに複数の機関が存在(4-1-3)	最長マッチング処理(全単語列の逐次サーチ)	4-3-2 4-5-1
同一の英語機関名を持つ機関が存在(4-1-4) －継承関係あり	機関名辞書の移行年と論文の発行年の比較(パターンに応じ条件を設定)	4-5-2
同一の英語機関名を持つ機関が存在(4-1-4) －継承関係なし	郵便番号マッチングの適用など	4-5-2
英語名が非常に似ている機関が存在(4-1-5) －ありふれた単語からなる名称	機関名辞書への別名の増強	4-4 4-5-3
英語名が非常に似ている機関が存在(4-1-5) －似た名称の大学	特別措置大学用統計辞書の利用	4-5-3
大学の下部組織同定の問題点(4-2)	・アドレス情報全体を対象としたマッチング処理 ・下部組織統計辞書の利用 ・機関名辞書補助セットの対象拡張 ・大学名と下部組織名の反転によるマッチング	4-6

5 論文の機関同定の結果 —大学下部組織同定を中心に—

これまでに述べた改善措置を2014年度に行った後、2015年5月に、WoSCCに1996-2013年に収録された日本所属機関データに対し機関同定処理を行った¹²。ここでは、32大学の下部組織同定を中心にその結果を述べるとともに、目視処理による検証で発見された問題点を併せて示す。

5-1 全体的な結果

5-1-1 機関同定できたデータの比率

機関同定結果の概要を図表16に示す。この表は、1996-2013年の18年間を6年ずつの3つの時期に区切り、機関同定されたデータ(セクター別に内訳を示す)と機関同定されなかったデータの出現頻度を示したものである。4-3-4で述べたように、機関同定されなかったデータには、セクターが推定される場合はセクター分類を与え、病院であることが推定される場合には病院フラグを付与するが、図表16にはそれらの出現度数も示している。なお、この表の数字は整数カウントによる出現機関数なので、1論文の中に複数の出現機関データがあればそれぞれが1件として数えられる。1論文の中に同一機関の異なる下部組織のデータが同定された場合も重複してカウントされる。

図表 16 機関同定結果の概要

	セクター	1996-2001	2002-2007	2008-2013	全期間
機 関 同 定	1 国立大学	351,325	440,229	485,315	1,276,869
	2 国立短大	132	90	2	224
	3 国立高専	2,755	3,193	3,136	9,084
	4 公立大学	33,287	42,334	45,443	121,064
	5 公立短大	443	281	123	847
	6 公立高専	204	310	279	793
	7 大学共同利用機関	9,802	13,546	13,861	37,209
	8 国の機関	35,275	15,106	12,322	62,703
	9 特殊法人・独立行政法人	36,795	102,854	124,479	264,128
	10 地方公共団体の機関	14,252	17,256	21,791	53,299
	12 私立大学	109,164	141,533	176,635	427,332
	13 私立短大	960	802	583	2,345
	14 私立高専	19	24	31	74
	15 会社	66,365	58,262	49,683	174,310
	16 非営利団体	16,995	20,792	21,799	59,586
	17 その他	167	214	187	568
		小計(A)	677,940	856,826	955,669
機 関 同 定 不 能	セクター同定	12,608	14,414	16,693	43,715
	病院同定	18,512	23,297	32,583	74,392
	同定不能	12,803	15,614	17,181	45,598
	小計(B)	43,923	53,325	66,457	163,705
	総計(C=A+B)	721,863	910,151	1,022,126	2,654,140
	機関同定率(%A/C)	93.9	94.1	93.5	93.8

¹² 機関同定処理は1981-2013年のデータに対して行ったが、機関名辞書は従来1996年以降の機関変遷を考慮しているため、ここでは1996-2013年のデータに対する結果を述べる。しかし、1981-95年のデータに対しても、少なくとも代表機関レベルでは、それ以降のデータと同程度の機関同定率を得ることができたことを確認している。

機関同定されたデータはどの時期もほぼ 94%である。同定できなかった 6%のデータには、機関名辞書に未収録の機関によるもの、機関名辞書に収録されている名称と異なる表記によるものがあり、これらのうちある程度出現頻度が高いものは辞書へのデータ追加を行っている。しかしこれ以外に、表記が不備あるいは曖昧で判断しづらい事例が相当数存在する。このようなデータは、無理に同定機関を決めると誤同定になる可能性があるため、むしろ同定不能のままにしておくことが望ましい。従って、機関同定率は 95%程度が限界ではないかと考えられる。

5-1-2 セクター別のシェア¹³

図表 17 に、機関同定されたデータの中での各セクターの占有率(シェア)を示す。国立、公立、私立の大学を合わせた占有率は 73~74%で安定しているが、その中では私立大学の比重が少しずつ増えている。1996-2001 年と 2002-2007 年の 2 つの時期の間に、国の機関のシェアが大きく減り、特殊法人・独立行政法人のシェアが急増しているのは、2001 年頃に国立試験研究機関から独立行政法人への移行が進んだことによる。しかし、特殊法人・独立行政法人のシェアの増加は、それだけでは説明できないほど大きい。他方、1996-2001 年に 10%近くあった会社の寄与は、2008-2013 年には 5%あまりまで減少している。

図表 17 機関同定されたデータのセクター別占有率(%)

セクター	1996-2001	2002-2007	2008-2013	全期間
1 国立大学	51.82	51.38	50.78	51.27
2 国立短大	0.02	0.01	0.00	0.01
3 国立高専	0.41	0.37	0.33	0.36
4 公立大学	4.91	4.94	4.76	4.86
5 公立短大	0.07	0.03	0.01	0.03
6 公立高専	0.03	0.04	0.03	0.03
7 大学共同利用機関	1.45	1.58	1.45	1.49
8 国の機関	5.20	1.76	1.29	2.52
9 特殊法人・独立行政法人	5.43	12.00	13.03	10.61
10 地方公共団体の機関	2.10	2.01	2.28	2.14
12 私立大学	16.10	16.52	18.48	17.16
13 私立短大	0.14	0.09	0.06	0.09
14 私立高専	0.00	0.00	0.00	0.00
15 会社	9.79	6.80	5.20	7.00
16 非営利団体	2.51	2.43	2.28	2.39
17 その他	0.02	0.02	0.02	0.02

5-1-3 マッチング処理ごとの同定率

4-3-2 から 4-3-3 で述べたように、機関同定は最長マッチング、曖昧マッチング、郵便番号マッチングの順に行われるが、どの時点で同定がなされたかを示すのが図表 18 である。98%以上が最長マッチング

¹³ セクター別の論文数の詳細な分析については、科学研究のベンチマーキング 2015, 調査資料-239, 科学技術・学術政策研究所, 2015 年 8 月(<http://www.nistep.go.jp/benchmark>)[5]を参照のこと。

処理で同定されており、この処理の重要性が判る。この表のマッチング種別コードCは、機械同定後の目視処理により同定を修正されたデータであり、5-3 で詳しく述べる。

図表 18 機関同定されたデータのマッチング種別

セクター	マッチング種別 *					計
	L	M	U	P	C	
1 国立大学	1,270,478	161	575	4,889	766	1,276,869
2 国立短大	218	0	6	0	0	224
3 国立高専	8,870	58	34	111	11	9,084
4 公立大学	119,488	64	73	1,053	386	121,064
5 公立短大	832	0	9	6	0	847
6 公立高専	787	1	0	5	0	793
7 大学共同利用機関	36,515	0	27	253	414	37,209
8 国の機関	61,611	16	300	220	556	62,703
9 特殊法人・独立行政法人	256,713	85	2,152	2,276	2,902	264,128
10 地方公共団体の機関	46,089	41	1,045	3,196	2,928	53,299
12 私立大学	420,800	453	1,010	3,168	1,901	427,332
13 私立短大	2,156	0	110	5	74	2,345
14 私立高専	56	1	1	16	0	74
15 会社	166,046	0	7,127	0	1,137	174,310
16 非営利団体	53,566	18	546	3,575	1,881	59,586
17 その他	566	0	2	0	0	568
合計	2,444,791	898	13,017	18,773	12,956	2,490,435
比率(%)	98.17	0.04	0.52	0.75	0.52	100.00

* マッチング種別コードの意味

コード	内容
L	最長マッチングによる同定
M	曖昧マッチングと郵便番号マッチングの結果が一致
U	曖昧マッチングによる同定
P	郵便番号マッチングによる同定
C	目視チェックにより同定を修正

5-2 32 大学の下部組織同定

ここでは、機関名辞書に下部組織情報を網羅的に収録している32大学について、下部組織同定の結果を述べる。

5-2-1 下部組織同定率の変化

図表 19 に、3 つの時期ごとの各大学の下部組織同定率を示す。全大学で見ると、時期が最近になるにつれて下部組織同定率は 85.9%→88.8%→91.8%と上昇している。機関名辞書には最新の下部組織(2013 年度以降)はほぼ漏れなく収録しているが、それ以前に存在していた下部組織は網羅されていないので、この結果は当然であるが、古い時期についてもかなりの高い割合で下部組織同定がなされている。下部組織の変遷が激しいことから考えるとやや意外な結果であるが、論文数の多い主要下部組織は比較的存在が安定しており、かつそれらの非現存下部組織の辞書収録率も高いためと考えられる。

図表 19 32 大学の出現度数と下部組織同定率

大学	1996-2001			2002-2007			2008-2013			全期間		
	出現度数	うち下部 組織同定	下部組織 同定率(%)	出現度数	うち下部 組織同定	下部組織 同定率(%)	出現度数	うち下部 組織同定	下部組織 同定率(%)	出現度数	うち下部 組織同定	下部組織 同定率(%)
北海道大学	16,014	14,184	88.6	20,922	19,244	92.0	23,051	21,654	93.9	59,987	55,082	91.8
東北大学	23,242	19,397	83.5	30,923	28,779	93.1	34,607	32,863	95.0	88,772	81,039	91.3
筑波大学	10,353	8,170	78.9	12,722	10,350	81.4	13,383	11,697	87.4	36,458	30,217	82.9
群馬大学	4,813	4,529	94.1	4,738	4,467	94.3	4,704	4,399	93.5	14,255	13,395	94.0
千葉大学	7,104	6,522	91.8	9,644	8,673	89.9	9,825	9,162	93.3	26,573	24,357	91.7
東京大学	40,550	37,509	92.5	51,093	47,651	93.3	58,748	55,487	94.4	150,391	140,647	93.5
東京医科歯科大学	4,693	2,257	48.1	5,999	2,357	39.3	7,682	3,441	44.8	18,374	8,055	43.8
東京工業大学	12,918	9,163	70.9	16,346	14,092	86.2	16,202	14,434	89.1	45,466	37,689	82.9
東京農工大学	2,762	2,033	73.6	3,831	2,833	73.9	4,686	3,672	78.4	11,279	8,538	75.7
新潟大学	5,779	5,210	90.2	7,327	6,629	90.5	7,227	6,619	91.6	20,333	18,458	90.8
富山大学	3,782	1,110	29.3	4,517	1,937	42.9	4,884	4,344	88.9	13,183	7,391	56.1
金沢大学	5,545	5,042	90.9	6,671	6,108	91.6	7,486	7,006	93.6	19,702	18,156	92.2
信州大学	4,659	4,446	95.4	5,399	5,077	94.0	6,213	5,785	93.1	16,271	15,308	94.1
岐阜大学	4,289	3,946	92.0	5,044	4,686	92.9	5,732	5,362	93.5	15,065	13,994	92.9
名古屋大学	17,101	14,616	85.5	20,650	18,034	87.3	22,891	21,191	92.6	60,642	53,841	88.8
京都大学	29,138	25,755	88.4	37,060	34,058	91.9	42,210	39,834	94.4	108,408	99,647	91.9
大阪大学	25,963	24,282	93.5	32,087	30,168	94.0	34,206	32,704	95.6	92,256	87,154	94.5
神戸大学	6,338	5,585	88.1	8,652	7,824	90.4	10,216	9,604	94.0	25,206	23,013	91.3
岡山大学	7,589	6,901	90.9	9,745	8,987	92.2	11,011	10,195	92.6	28,345	26,083	92.0
広島大学	9,393	7,127	75.9	12,020	10,175	84.7	12,228	11,065	90.5	33,641	28,367	84.3
徳島大学	4,493	3,969	88.3	5,143	4,460	86.7	5,389	4,837	89.8	15,025	13,266	88.3
九州大学	17,348	15,521	89.5	21,732	19,652	90.4	25,517	23,581	92.4	64,597	58,754	91.0
長崎大学	4,925	4,489	91.1	6,146	5,687	92.5	6,757	6,227	92.2	17,828	16,403	92.0
熊本大学	5,172	4,416	85.4	5,651	4,980	88.1	6,848	6,265	91.5	17,671	15,661	88.6
大阪市立大学	5,151	4,835	93.9	6,033	5,448	90.3	5,534	4,858	87.8	16,718	15,141	90.6
大阪府立大学	3,679	1,846	50.2	4,173	2,817	67.5	4,148	3,614	87.1	12,000	8,277	69.0
慶應義塾大学	7,088	6,324	89.2	9,569	8,502	88.8	12,591	11,705	93.0	29,248	26,531	90.7
東海大学	3,866	2,998	77.5	4,450	3,491	78.4	4,631	4,007	86.5	12,947	10,496	81.1
東京理科大学	4,381	3,815	87.1	5,170	4,499	87.0	5,726	5,209	91.0	15,277	13,523	88.5
日本大学	3,972	3,688	92.8	6,761	6,363	94.1	8,239	7,863	95.4	18,972	17,914	94.4
早稲田大学	3,792	2,954	77.9	5,855	4,673	79.8	7,714	6,211	80.5	17,361	13,838	79.7
近畿大学	3,097	2,678	86.5	3,796	3,395	89.4	5,350	5,015	93.7	12,243	11,088	90.6
総計	308,989	265,317	85.9	389,869	346,096	88.8	435,636	399,910	91.8	1,134,494	1,011,323	89.1

5-2-2 下部組織同定に対する補助辞書の貢献

マッチング処理で 32 大学のいずれかが代表機関として同定されると、次の手順で下部組織同定がなされる。

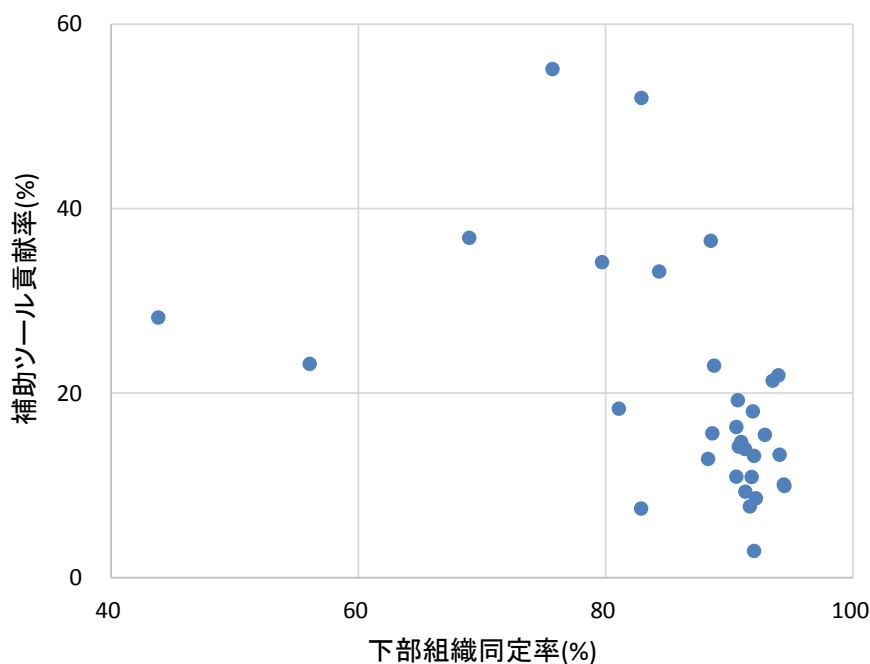
- ① 機関名辞書とのマッチング
- ② 拡張された補助セットとのマッチング(4-6-3 参照)
- ③ 下部組織統計辞書とのマッチング(4-6-2 参照)

各大学の下部組織同定データが①～③のどの時点でマッチしたかの内訳を図表 20 に示す。下部組織同定に対する①、②、③の貢献率はそれぞれ82%、7%、11%であり、補助ツールとして導入した②と③がかなりの効果を及ぼしたことが判る。図表 20 には、大学別に②と③を合わせた貢献率も示しているが、大学により 3%～55%と大きな差がある。この数字と下部組織同定率には相反関係があることが、図表 21 から見て取れる(両者の間の Pearson 相関係数は-0.46)。すなわち、これらの補助ツールはもともと下部組織同定率が低い大学でより大きな効果を発揮しており、大学間の下部組織同定率の格差縮小に寄与している。

図表 20 下部組織マッチングがなされたツール別内訳(%)

大学	①機関名辞書	②拡張補助セット	③下部組織統計辞書	②+③
北海道大学	89.1	8.1	2.8	10.9
東北大学	86.1	3.4	10.5	13.9
筑波大学	92.5	5.8	1.7	7.5
群馬大学	78.1	1.4	20.5	21.9
千葉大学	92.3	3.9	3.8	7.7
東京大学	78.7	3.2	18.2	21.3
東京医科歯科大学	71.8	0.0	28.2	28.2
東京工業大学	48.0	27.2	24.8	52.0
東京農工大学	44.9	35.6	19.5	55.1
新潟大学	85.8	3.8	10.4	14.2
富山大学	76.8	18.6	4.6	23.2
金沢大学	91.4	5.0	3.5	8.6
信州大学	86.7	2.4	10.9	13.3
岐阜大学	84.6	5.2	10.3	15.4
名古屋大学	77.0	8.4	14.5	23.0
京都大学	82.0	6.4	11.6	18.0
大阪大学	90.1	1.1	8.8	9.9
神戸大学	90.7	7.9	1.4	9.3
岡山大学	86.8	4.6	8.6	13.2
広島大学	66.8	4.7	28.5	33.2
徳島大学	87.1	2.1	10.8	12.9
九州大学	85.3	10.2	4.5	14.7
長崎大学	97.1	1.1	1.8	2.9
熊本大学	84.4	9.3	6.3	15.6
大阪市立大学	89.1	3.8	7.1	10.9
大阪府立大学	63.2	11.7	25.1	36.8
慶應義塾大学	80.8	0.7	18.5	19.2
東海大学	81.7	14.6	3.7	18.3
東京理科大学	63.5	18.5	18.0	36.5
日本大学	89.9	3.8	6.3	10.1
早稲田大学	65.8	26.5	7.7	34.2
近畿大学	83.7	12.0	4.3	16.3
全体	81.9	6.7	11.4	18.1

図表 21 下部組織同定率と補助ツール貢献率の関係



5-2-3 下部組織同定がなされない理由

下部組織同定がされない(つまり代表機関である大学に同定される)データが全体的には 10~15%存在するが、その理由は次の 2 つに大別される。

- (1) 所属機関データに大学名(もしくは大学名+キャンパス名)しか書かれていない。
- (2) 表記されている下部組織名が機関名辞書や補助ツールとマッチしない。

このことを図表 22 で説明する。ここでは、3 つの時期すべてで下部組織同定率が 85%以上の 22 大学(図表 22(a))とそれ以外の 10 大学(図表 22(b))を別の図に示す。目視確認から、図表 22(a)に示した大学では、下部組織同定されなかったデータの多くが(1)に属することが分かっている。言い換えると、大学のデータの 5~15%はこの種の表記がなされている。これらのデータから下部組織同定をすることは不可能なので、現状では 90%前後が下部組織同定率の限界と考えられる。なお、図表 22(a)中で大阪市立大学のみ下部組織同定率が低下している(1996-2001 年の 94%から 2008-2013 年の 88%へ)が、上記(1)のデータが増えているためと考えられる(2008-2013 年の下部組織同定不能データ中 78%がこのようなデータであった)。

一方、図表 22(b)に示した大学では、下部組織が表記されていても、下部組織同定が行われていない場合も多く、(1)とともに(2)のデータが相当混じっていると考えられるが、そのいくつかについて個別に考察する。

まず、東京医科歯科大学は、他の大学に比べて下部組織同定率がきわめて低く、どの時期も 50%に達

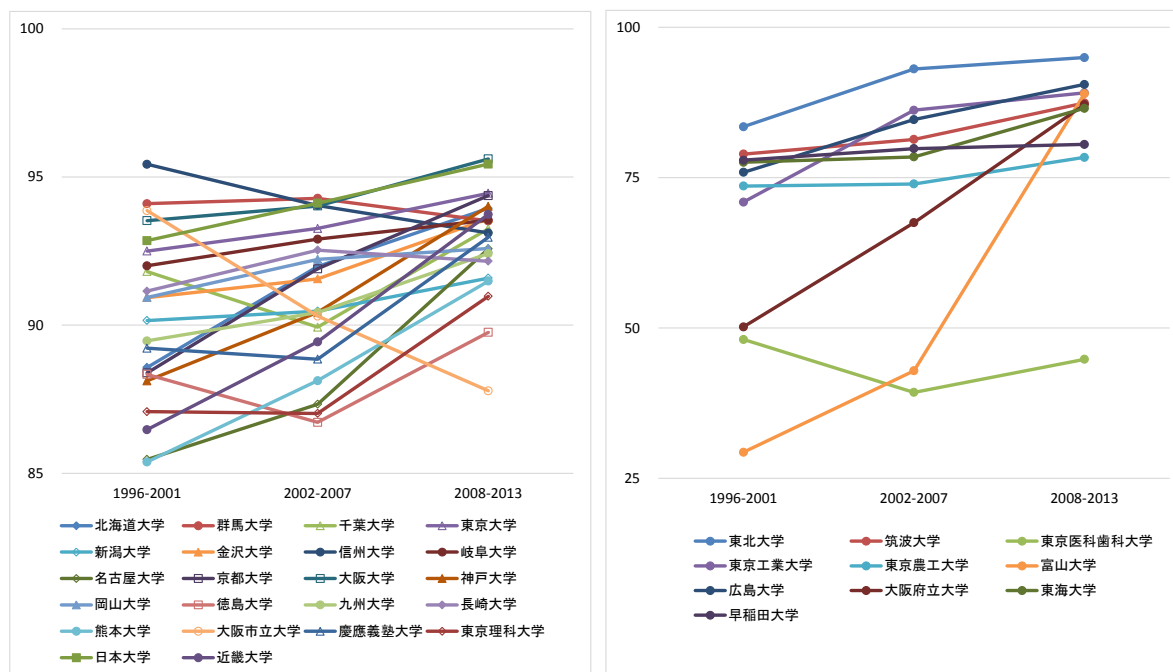
しない。この大学は、学部では医学部と歯学部、大学院では医歯学総合研究科が規模において大部分を占めているため、所属機関表記で第 2 階層下部組織に対応する学部名、研究科名が省略されて、その下の第 3 階層下部組織に対応する学科名等が直接表記されることが非常に多い。学部名や研究科名が書かれたデータが少ないと、下部組織統計辞書を作るのに必要な統計的信頼性を持ったデータが得られないことになる。従ってこの大学については、現状では下部組織レベルでのデータ分析を行うことは難しい。また、単科大学に近い形態であることから、他の 31 大学に比べて、下部組織レベルの分析を行う必要性も低いと考えられる。

東京医科歯科大学ほどではないが、東京農工大学、早稲田大学の下部組織同定率もあまり高くない。3 つの時期の同定率は、前者が 75%前後、後者が 80%前後である。両大学とも、その大きな理由は、下部組織同定ができなかったデータのうち上記(1)のデータの割合が高いことで、2008-2013 年において、東京農工大学では 60%、早稲田大学では 66%に上っている。なお、東京農工大学は、機関名辞書とのマッチングの段階での下部組織同定率は東京医科歯科大学と同じくらい低い(30%強)が、拡張補助セットと多くのマッチングがなされたために、下部組織同定率に大きな差が付いた(図表 20 参照)。これは、東京農工大学の主要第 2 階層組織が農学部(及び農学府)と工学部(及び工学府)であるため、第 3 階層下部組織の名称しか書かれていなくても、それが属する第 2 階層組織が比較的容易に判別されるためである。

図表 22 下部組織同定率の変遷

(a) 全時期で同定率 85%以上の 22 大学

(b) それ以外の 10 大学



富山大学と大阪府立大学は、3 つの時期の間で下部組織同定率がそれぞれ 29.3%→42.9%→88.9%、

50.2%→67.5%→87.1%と急上昇しているが、これには両大学において他大学との統合があったことが影響している。富山大学は2005年に富山医科薬科大学及び高岡短期大学と、大阪府立大学は2005年に大阪女子大学、2009年に大阪府立看護大学と統合した。ここでの集計にはこれら統合された大学のデータも含んでいるが、それらの統合前の大学の同定データは下部組織同定に数えられていない。従って、統合された旧大学のデータを除いて計算すれば、古い時期の下部組織同定率はずっと上昇すると考えられる。

図表 22(b)にあるこれ以外の大学(東北大学、筑波大学、東京工業大学、広島大学、東海大学)は、いずれも下部組織同定率が徐々に上昇し、2008-2013年には87%以上に達している。

東京医科歯科大学を除く31大学では、2008-2013年における下部組織同定率は75%を越えているので、各種分析に利用する際のベースラインとしての下部組織同定はなされていると考えられる。

5-3 目視による同定結果の検証

機械的に同定された結果から問題がありそうなパターンを取り出してチェックを行い、延べ約13,000件のデータの同定先を修正した。図表 18 のマッチング種別コード C がそれで、同定されたデータの約0.5%に当たる。ここでは、抽出された主な問題点とその修正処置について示す。これらの修正は取りあえず手作業で行ったが、今年度に同定アルゴリズムあるいは辞書類に反映して改善の予定である。

5-3-1 特別措置大学

英語名称がよく似ているため誤同定が起りやすい大学の組み合わせ(図表 13 参照)に対する特別措置の方法を4-5-3に述べた。しかし、統計辞書による同定が正しくないケースが見られ、その中には、統計辞書作成のための参照データとして用いたデータベース中の郵便番号が間違っていることが原因のものもあった。特に、金沢大学と金沢医科大学の場合、両方に医学部があるためこのような間違いが多く見られた。

これらの誤りは手作業で修正したが、今後特別措置大学用統計辞書(4-5-3 参照)を見直し、特徴的な学部や学科の名称で判別するように改善する必要がある。たとえば、岩手大学と岩手医科大学の場合、Sch Med のような医学部を示す語があれば岩手医科大学と判断してよい。他方、金沢大学と金沢医科大学の場合、医学部関係の語は統計辞書の対象から除くことが妥当である。

5-3-2 単科大学

教育大学、薬科大学、工学系の単科大学で、英語名が「YYYY University XXXX」となっていた場合、XXXXの記述が辞書とマッチしないとYYYY大学にマッチしてしまうものが見られた。たとえば、北海道教育大学(Hokkaido University of Education)がHokkaido Univ Edと表記されていると、これは辞書にない

め北海道大学に同定される(Educ なら略記辞書にあるため正しく同定される)。これらは手作業で修正したが、多くは XXXX 部分の略記を機関名辞書か略記辞書に追加することにより対応可能と思われる。

5-3-3 連合大学院

連合大学院には名称に United Graduate School が付くが、それがなくて参加している大学が羅列されていると、それらの個別の大学とマッチングが行われ、複数の大学に同定されてしまう。連合大学院は代表機関の下部組織としてのみ扱うこととし、上記のように複数の大学へのマッチングが行われている場合、ORG サブフィールドに記述されている大学に対するマッチング結果のみを残した。

5-3-4 複数の同定候補から 1 つを選び出す際の問題

現在のアルゴリズムでは、マッチングにより複数の同定候補が得られた場合、それらをペアごとに比較し、優先順位を定めていずれかを抽出するようになっている。しかしながら、A, B, C の 3 つの同定候補が得られた場合、A よりは B、B よりは C、C よりは A のように、優先順位が決まらないためどの機関に同定するか判別できず、結果が出力されないことがある。これらについて手作業で修正した。特に、独立行政法人農業・食品産業技術総合研究機構(NARO)は多くの下部組織研究所や前身の機関を持つため、この事例に相当する関連データが多数あった。

5-3-5 病院を決める基準

現在の機関名辞書には、明らかに病院にもかかわらず病院フラグが付与されていない機関が存在する(病院フラグを補助セットに収録しているのは信頼性が十分でないことによる)。何らかのオーソリティ・リストに従って病院を判定することを今後検討したい。

5-3-6 複合同定にかかわるもの

一つの所属機関フィールドに複数の機関が存在する場合とその処理(複合同定)についてはそれぞれ 4-1-3、4-5-1 で述べたが、必ずしもその処理が適当でない場合もある。出現回数が 3 件以上の複合同定を目視で確認した。次のようなパターンが見出された。

- (1) 国立試験研究機関が英語名称に省庁名と研究所名を両方含んでいると、省庁と研究所の両方に同定される。たとえば National Food Research Institute, MAFF という表記は、農林水産省と農林水産省食品総合研究所に複合同定される。この場合は研究所への同定が正しいので、省庁への同定については削除した。このような事例については、同定処理時の特別処理が必要と考えられる。
- (2) 赤十字の病院で、認可法人日本赤十字と独立行政法人国立病院機構の病院に複合同定されるものが散見された。認可法人日本赤十字社大阪赤十字病院が、正式名称である Osaka Red

Cross Hospital と表記されていればこのようなことは起こらないが、Red Cross Osaka Hospital と表記されると、その一部が独立行政法人国立病院機構大阪医療センターにもマッチしてしまう。これらは、認可法人日本赤十字の病院のみを残した。

- (3) 機関名辞書に登録されていない機関が、その名称によって複合同定されるケースが見られた。例として、「オリンパステルモバイオマテリアル株式会社」のように複数の会社の名称を組合せたような名称の機関(辞書には登録されていない)が、株式会社バイオマティクス、テルモ株式会社、オリンパス株式会社のように複数の会社に同定される。この場合機関同定を取り消し、セクター分類を会社とした。別の例で、「NTT 西日本大阪病院」が独立行政法人国立病院機構大阪医療センターと西日本電信電話株式会社に複合同定されたので、これも機関同定を削除し、病院フラグを付与した。
- (4) 継承関係もないのに英語名がたまたま同名になる場合の処理は 4-5-2 の(2)に示したが、それでも複合同定になることがある。清泉女子大学と聖泉大学(どちらも英語名は Seisen University)はその例である。これらは住所から判定し、もっともらしい方の大学を残した。
- (5) 会社名が短い場合(株式会社クボタ商会、株式会社クボタなど)、複数の会社が同定される場合が見られた。これについては、研究開発をしていると思われる会社を残した。

5-4 機関名の同定精度の検証

NISTEP 論文機関名同定プログラム(Web of Science バージョン)による機関名の同定及び目視による確認・修正を経た、最終的なデータセットにおける機関名の同定精度の検証結果を図表 23 に示した。

ここでは、1999-2013 年までの 15 年分について、年毎に 200 アドレスをランダムサンプリングし、それらについて目視による確認を行った。なお、一つのアドレスに対して複数の機関が同定されている場合があるので、実際のサンプル数は 3,000 よりも多くなっている。目視確認した 3,014 サンプルの内、2,970 については機関名(もしくはセクター又は病院フラグの同定)が正しくなされており、正解率は 98.5%である。

誤同定の 14 件(0.5%)については、企業名の誤同定が多く、セクター判定には影響はない。未同定の 30 件のうち 23 件(0.8%)の多くは、アドレス情報が十分でなく、プログラムでは同定できなかったが、人が Web 上の検索等を駆使すれば、機関名のあたりをつけることはできるというケースであり、非営利団体、組織の略称名、現存しない機関などである。他の 7 件については、機関同定を行なうための情報が欠損しているレコードであった。

図表 23 最終的なデータセットにおける機関名の同定精度の検証結果

	サンプル数	同定	誤同定	未同定
該当数	3014	2970	14	30
割合	100.0%	98.5%	0.5%	1.0%

6 まとめと今後の課題

本報告書では、NISTEPにおける機関名辞書の整備(第2, 第3章)と、その活用例としての論文データベース中の著者所属機関同定(第4, 第5章)について述べた。特に、32大学における下部組織情報の機関名辞書への収録と、その結果可能になった下部組織レベル同定に焦点を当てた。

要約すると以下のようなになる。

第2章では、研究開発に関する動向把握や方針・政策検討の基礎データとして研究機関の基本情報をまとめた辞書の必要性、特に変遷情報や主要下部組織情報のアーカイブ化の重要性を述べた。その上で、機関名辞書の構成とデータ収集方法について概説した。

第3章では、大学の下部組織レベルの情報の重要性に鑑み、論文数の多い32大学の下部組織情報を網羅的に機関名辞書に収録することにしたことを述べ、それらの情報の収集方法を示した。その後、最近の大学の組織構造が多様化、複雑化していること、大学 Web サイトからの下部組織情報入手に困難があることを指摘した。

第4章では、機関名辞書を用いた論文データベースの機関同定について述べた。機関データ表記の多様性による同定上の種々の問題点を挙げた後、機関同定アルゴリズムの骨子及び同定精度の向上(大学下部組織の同定率上昇を含む)のための措置(機関名辞書のデータ増強、補助ツールの開発、同定アルゴリズムの工夫等)について論じた。また、第5章では、この方法をWoSCCデータベースに適用した同定の結果についても簡単に述べた。

以上に述べた取り組みを踏まえ、機関名辞書の継続的維持と大学下部組織情報の拡張の2点が、今後の課題として重要と考えられる。

機関名辞書が研究開発動向の把握と分析のための基礎データとしての意義を持つとすれば、第1の課題の重要性は当然であろう。機関やその下部組織は常に変遷しており、その情報を確実にとらえてアーカイブ化するためには、不断の状況把握・情報更新が必要である。以下の方法による情報収集とデータ更新を行う体制と予算措置を継続的に講じられるよう検討を進めたい。

- 大学、短大、高専、大学共同利用機関、国の機関、特殊法人・独立行政法人、学校法人の各セクターに属する代表機関については、毎年少なくとも1回、Webサイトのリスト(現在は図表4に示すものが存在)を調べる。
- これらの機関の下部組織で機関名辞書のコアセットに含まれるもの(2-3参照)については、少なくとも年1回、それぞれが属する代表機関のWebサイトで存否を確認する。
- 地方公共団体の機関及び非営利団体については、数年に1回程度を目標に、辞書に含まれる機関についてWeb調査を行い、機関の変遷、英語名の変更等の有無を確認する。
- 民間企業については、NISTEP企業名辞書との連携を進め、年に一度、必要な情報(特に変遷情報、所在地情報)を移入する。

- 新設等により辞書に新たに登録された機関については、その機関の Web サイトから、正式名称（日本語と英語）と所在地の情報を入手する。

第 2 の課題について、現在下部組織情報の網羅的収録の対象としている 32 大学は、1996～2010 年の期間の論文数を基準として選定している(3-1 参照)。しかし、その実態は年とともに変化するので適宜見直しが必要である。本来は、ある程度の研究規模を持つ総合大学をすべて対象とするのが望ましい。しかし、第 3 章で述べたように、現在の 32 大学だけでも、毎年各大学の下部組織の変遷を調査することはかなりの負担である。この作業を軽減するためには、文部科学省の関係課室あるいは個々の大学との連携により、毎年確実に各大学の組織変遷情報を知ることが望ましい。どの方法が適切か、今後検討したい。

参考文献

- [1] “データ・情報基盤－政策研究の高度化とエビデンスベースの政策形成のためのツール．” 科学技術・学術政策研究所. <http://www.nistep.go.jp/research/scisip/data-and-information-infrastructure>
- [2] 小野寺夏生. 大学・公的機関における研究開発に関するデータの整備－マイクロデータ分析への貢献－. NISTEP NOTE No.11. 科学技術・学術政策研究所, 2014年5月. <http://hdl.handle.net/11035/2926>
- [3] 全国試験研究機関名鑑 2008-2009. 文部科学省科学技術・学術政策局監修, 丸善, 2008.
- [4] 全国大学職員録 平成10年版(国公立大学編, 私立大学編). 廣潤社, 1998.
- [5] 阪 彩香, 伊神 正貴. 科学研究のベンチマーキング 2015. 調査資料-239. 科学技術・学術政策研究所, 2015年8月. <http://www.nistep.go.jp/benchmark>

調査担当

本報告書で述べられている結果は、文部科学省の「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」の一環として実施している「大学・公的機関における研究開発に関するデータ整備」の成果である。

調査の全体設計、データの集計等は文部科学省科学技術・学術政策研究所が担当した。NISTEP 大学・公的機関名辞書の整備、NISTEP 論文機関名同定プログラム(Web of Science バージョン)の開発及びそれを用いた機関名同定作業は株式会社 RNAi が担当した。

文部科学省科学技術・学術政策研究所

(調査実施、報告書執筆)

小野寺 夏生 科学技術・学術基盤調査研究室 客員研究官

[調査の全体とりまとめ、NISTEP 大学・公的機関名辞書全般の設計・整備・検証、NISTEP 論文機関名同定プログラム(Web of Science バージョン)全般の開発及びそれを用いた機関名同定作業の検証、報告書執筆]

(調査実施、報告書執筆協力)

伊神 正貫 科学技術・学術基盤調査研究室長

[NISTEP 論文機関名同定プログラム(Web of Science バージョン)を用いた下部組織同定率向上策の検討、機関名同定結果の確認、機関名の同定精度の検証、報告書執筆協力]

(調査実施)

阪 彩香 科学技術・学術基盤調査研究室 主任研究官

[NISTEP 論文機関名同定プログラム(Web of Science バージョン)を用いた下部組織同定率向上策の検討、機関名同定結果の確認及びそれに基づく同定結果の改善、機関名の同定精度の検証]

(調査補助)

若林 真知子 科学技術・学術基盤調査研究室 事務補助員(2015年3月末まで)

[32大学についての下部組織情報の取得及び取得した情報についての確認作業の実施]

株式会社 RNAi

坂本 義弘、山田 智之

[NISTEP 大学・公的機関名辞書の整備、NISTEP 論文機関名同定プログラム(Web of Science バージョン)の開発及びそれを用いた機関名同定作業]

(2015年9月末時点)

(裏白紙)

NISTEP NOTE (政策のための科学) No. 15

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—

2015年10月

文部科学省 科学技術・学術政策研究所
科学技術・学術基盤調査研究室
〒100-0013

東京都千代田区霞が関3-2-2 中央合同庁舎第7号館 東館16階
TEL:03-6733-4910 FAX:03-3503-3996