

Title	科学技術イノベーション政策立案のためのデータプラットフォーム：投資と成果のデータ対応調査
Author(s)	原田，裕明；池内，健太；原，泰史；黄，俊揚；黒田，昌裕
Citation	年次学術大会講演要旨集，31：620-623
Issue Date	2016-11-05
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/13856
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

科学技術イノベーション政策立案のためのデータプラットフォーム ー投資と成果のデータ対応調査ー

○原田 裕明(JST/CRDS^{*1}), 池内 健太, 原 泰史, 黄 俊揚(GRIPS^{*2}), 黒田 昌裕 (JST/CRDS^{*1})

(^{*1} 国立研究開発法人 科学技術振興機構 研究開発戦略センター)

(^{*2} 国立大学法人 政策研究大学院大学 SciREX センター)

1. 背景と課題

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」(SciREX)推進事業の一環として、エビデンス・データを基に、いくつかの政策オプションを合理的に導出する研究を進めている[1]。最近ではエビデンス・データとして論文や特許だけでなく、研究資金や人材のデータなども整備されつつあり、その利用機運が高まっている[2][3][4]。一般に、政策を検討する場合には、データの収集から分析するまでの期間が短いため、必要と考えられるエビデンス・データの集合(データベース)間をあらかじめ自在に相互参照できるように準備しておくことが必要となる。

また政策の論議には、たとえば「ビッグデータ」、「深層学習」など最先端の用語が出現することが多いが、なるべくタイムリーかつ掘り所が明確な用語集があることが望ましい。この要求に対応するために、JST/CRDS では2年おきに、幅広い科学技術分野について500名以上の識者の意見にもとづき、「研究開発の俯瞰報告書」¹を編集し、公開している。ここには最新の科学技術動向を代表するキーワードである「研究開発領域」(以下、俯瞰キーワードとよぶ)が計355件(2015年版)挙げられており、個々について用語説明、国際的な研究状況、関連用語などが2,700ページにわたって記載されている。最新で豊富な情報と根拠を持つ俯瞰キーワードを政策論議に利用する意義は大きい。

基礎となるエビデンス・データや俯瞰キーワードのような科学技術用語は整備されつつあるといえるが、なお解決すべき根本的な課題は次の2点であると考ええる。

(1) 科学技術用語の流動性

俯瞰キーワードのような科学技術用語の中にはタイムリーであるがゆえに寿命は短く、また定義が抽象的な用語も多い。したがって新しい用語を使って既存のデータベースを検索するためには、シソーラスなどの辞書を使用できる専門的な知識が新たに必要となる。

(2) 基礎データの増加と、科学技術用語との柔軟な対応付け

論文や特許などの基盤となるデータベースには過去から膨大な蓄積データがある上、毎年、新たな追加が生じる(たとえば特許出願データは年31万件ずつ追加されている[5])。追加分のデータ登録時に厳密な分類コード(IPCなど)が付けられるが、上記(1)のような新しい用語との対応付けは通常はおこなわれない。新しい用語が出現するたびに対応付けすることは量から見て不可能である。

◆二つの方策案

特に新しい用語特有の曖昧さや変化の速さ、検索の精度と処理時間のバランスなどを前提とすると、データベース間の相互参照の現実的な方法として次の2方策が考えられる(図1)。

方策(A)：間接的対応付け

コンコダンス表により、分類コード間の簡易な対応付けをおこなう。この表はキーワード x (=データベース P の分類コードの一つ) に対するデータベース Q の分類コードの近似性 $F(x)$ をあらかじめ記録したものである(図2)。 $F(x)$ が閾値 Th 以上の Q の分類コードのみ検索対象とすれば処理時間が少なくなる。また高い精度が必要ない場合には、その選択した分類コードに含まれる全データに一律に近似性 $F(x)$ の重みを付けて簡易計算することによって、さらなる高速化が望める。

方策(B)：直接的対応付け

分類コード経由ではなく、データベースのデータ内容にキーワードを含んでいるかを直接判定して検索する。全データに対する個別判定であるため、方策(A)に比べて処理時間がかかる。

¹ <http://www.jst.go.jp/crds/report/report02/index.html> にて公開。環境・エネルギー、ライフサイエンス・臨床医学、情報科学技術、ナノテクノロジー・材料、システム科学技術の5分野。

◆フィージビリティスタディの目標と評価尺度

今回は上記2つの方策を前提として、表1に示した4つのデータベース（科研費、論文、特許、研究開発の俯瞰報告書）の分類コードを対象としたフィージビリティスタディをおこなった。

このフィージビリティスタディの目標はまず方策(A)の実行可能性を探ることとし、コンコーダンス表によって検索の手間がどれだけ軽減されるかを評価尺度とした。すなわち精度、速度とも総当たり探索に比べて優位であれば方策(A)を採用し、そうでなければ方策(B)の可能性を探ることにした。

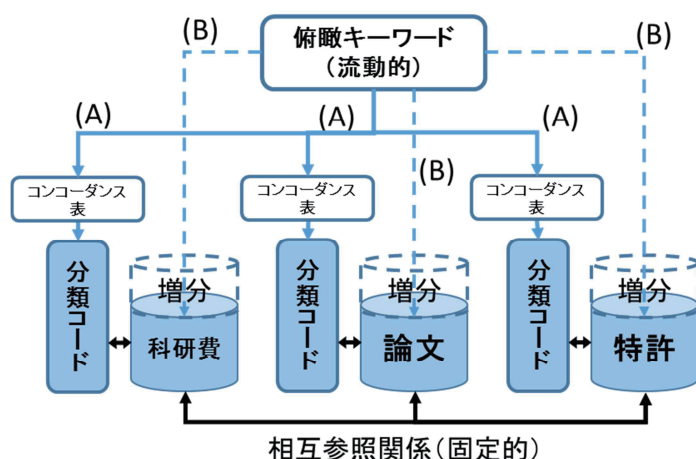


図1. データベース間の2つの参照方法

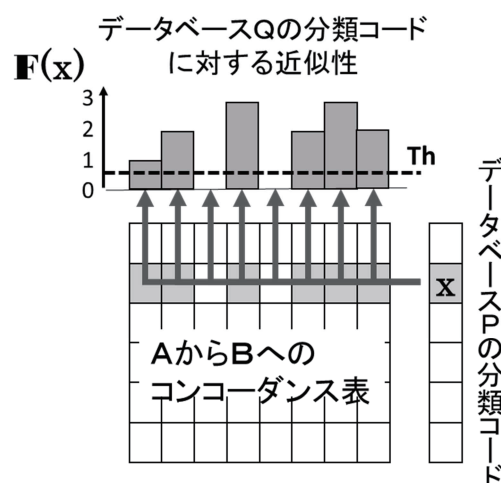


図2. コンコーダンス表の考え方

2. 実験

実験では俯瞰キーワードの中から「CPS/IoT」²領域のみ検索の対象とした。この領域を選択したのは、比較的新しいキーワードであり、複数の科学技術の分野にまたがっていると予想したためである。

表1. 検討対象としたデータベース

データベース名 (検索サイト)	検討対象とした 分類コード	実験範囲
科研費 (KAKEN)	分野・分科・細目・ キーワード	農学、総合理工、情報学、総合領域、複合領域、社会科学、数理系科学、化学、工学の中の関連173細目
論文 (J-GLOBAL)	JST 技術分類	システム・制御工学 (I)、情報工学 (J)、電気工学 (N)、機械工学 (Q) の4分野
特許 (J-PlatPat)	IPC サブクラス	全セクション (A 生活必需品～H 電気) で317サブクラス
俯瞰キーワード (研究開発の俯瞰報告書)	研究開発領域	CPS/IoTに関する6領域 (CPS/IoT アーキテクチャー、M2M、社会システムデザイン、CPS/IoT セキュリティ、応用と社会インパクト、ものづくりとIoT)

◆実験1（主観分類）

(ア) 俯瞰キーワード×科研費コード（表3）

俯瞰キーワード（6件）と科研費173細目との近似性を測った。この測定は各キーワードの説明などを手がかりとして複数の調査員による主観評価（4水準）でおこなった。

(イ) 俯瞰キーワード×JST 技術分類（表4）

「システム・制御工学 (I)」、「情報工学 (J)」、「電気工学 (N)」、「機械工学 (Q)」の4分野を選択して、俯瞰キーワードとの近似性を同様に主観評価によって測った。

◆実験2（内容検索）

(ウ) 俯瞰キーワードによる論文の検索件数（表5）

J-GLOBAL の検索機能を用いて俯瞰キーワードが論文のタイトル、抄録などの中に出現する論

² CPS: Cyber Physical System, IoT: Internet-of-Things

(エ) 俯瞰キーワードによる公開特許の検索件数 (表 6)

3. 検討

実験 1 の主観分類結果として、図 3 に俯瞰キーワードから科研費コードを参照する場合のコンコードダンス表を例示する。CPS/IoT の関連として「情報学」や「数学」への近似性が特に強く現れているのがわかる。また表 3、表 4 は各近似性の水準ごとに全体に対する比率をまとめたものである。表 3 では、6 つの俯瞰キーワードのうち 4 つで「まったく関係がない」水準の比率が 50%未満であった（網かけ部）。表 4 でも似た傾向が見られた。これは俯瞰キーワード自体の定義の曖昧性や、各種の分類コードとの粒度の違いなど、分類コードの本来的な原因に因るものと考えられる。結果的に、科研費や論文データベースに対して、これらの俯瞰キーワードから分類コード全体の 6~7 割以上も「何らかの関係がある」と見なして検索しなければならないことを意味する。

実験 2 は俯瞰キーワードがデータ内容に含まれているものを直接計数しているので、この計数の分布が分類コードに対する近似性をほぼ表している。反面、データを逐一検索するために多大の手間を要した他に、キーワードが一つ含まれるだけで同類であると単純判定している点は課題である。

方策	本来の長所	実験によって明らかになった問題点
方策(A) : 間接的対応付け	コンコーダンス表をいったん作成しておけば、俯瞰キーワードに対するデータの検索範囲を近似性によって効率よく絞ることができる。	近似性が「何らかの関係がある」ものを無視できない。そのため検索範囲が拡散して効率性の長所が生かせない。
方策(B) : 直接的対応付け	検索精度が高い。	検索の手間がかかる。精度向上のためにはより高度な近似性判定が必要である。

主観分類で作成したコンコダンス表の近似性精度をより高めて、検索範囲の拡散を防ぐ対策として、実験2のような直接的な検索結果を用いることが考えられる。精度の高いコンコダンス表を一度作成すれば検索効率が向上するが、作成の手間がかかると方策(A)の長所が発揮できない。すなわち方策(A)の優位性は直接検索(方策(B))の効率に依存しており、方策(A)よりも方策(B)の追求が先決となる。

今後は政策の議論のためにどの程度の精度と速度が要求されるかを調査・検討する必要がある。

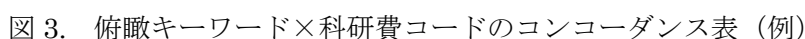


表 3. 俯瞰キーワード×科研費コード（主観による対応付け）の近似性分布

近似性の水準	CPS/IoT アーキテ クチャー	M2M	社会シス テムデザ イン	CPS/IoT セキュリ ティー	応用と社 会インパ クト	ものづく りと IoT
3 ほとんどが俯瞰キーワードに関係する	14%	14%	14%	14%	14%	14%
2 一部が俯瞰キーワードに関係する	12%	9%	8%	5%	12%	9%
1 俯瞰キーワードと無関係ではない	63%	43%	35%	2%	20%	39%
0 まったく関係がない	11%	33%	43%	78%	53%	38%

（たとえば俯瞰キーワード「M2M」に対して 173 細目中の 33%は「まったく関係ない」と分類された。）

表 4. 俯瞰キーワード×論文 JST 技術分類（主観による対応付け）の近似性分布

近似性の水準	CPS/IoT アーキテ クチャー	M2M	社会シス テムデザ イン	CPS/IoT セキュリ ティー	応用と社 会インパ クト	ものづく りと IoT
3 ほとんどが俯瞰キーワードに関係する	14%	12%	6%	14%	2%	6%
2 一部が俯瞰キーワードに関係する	43%	39%	20%	10%	31%	18%
1 俯瞰キーワードと無関係ではない	33%	20%	10%	4%	26%	55%
0 まったく関係がない	11%	29%	64%	72%	41%	21%

表 5. 俯瞰キーワードによる論文の検索件数

JST 分類コード	分類見出し	CPS/IoT アーキテ クチャー	M2M	社会シス テムデザ イン	CPS/IoT セキュリ ティー	応用と社 会インパ クト	ものづく りと IoT
JE04	図形・画像処理	40	6,351	2,278	1,066	16	939
JE04010I	図形・画像処理一般	38	5,210	1,776	1,019	16	520
JE04020T	医用画像処理	2	1,170	510	48	0	426

表 6. 俯瞰キーワードによる公開特許の検索件数

IPC 分類 セクション	IPC 分類 セクション内容	CPS/IoT アーキテ クチャー	M2M	社会シス テムデザ イン	CPS/IoT セキュリ ティー	応用と社 会インパ クト	ものづく りと IoT
A	生活必需品	1	12	1,214	188	62	20
B	処理操作・運輸	0	53	1,661	53	4	9
C	化学・冶金	1	56	802	14	2	0
D	繊維・紙	0	2	45	4	0	0
E	固定構造物	0	9	56	68	1	0
F	機械工学他	0	10	111	23	1	1
G	物理学	22	268	1,607	846	142	372
H	電気	71	379	1,241	481	56	59

【謝辞】本実験の実施を支援していただいた株式会社三菱総合研究所の土谷和之氏はじめチームの皆様
に感謝いたします。

【参考】

- [1] JST 研究開発戦略センター「科学技術イノベーション政策の科学における政策オプションの作成～ICT 分野の政策
オプション作成プロセス～」、CRDS-FY2015-RR-07（2016.3）
- [2] JST 研究開発戦略センター「変動の時代に対応する科学技術イノベーション政策のためのエビデンスの整備と活用
に向けて」、CRDS-FY2015-RR-01（2015.3）
- [3] 文部科学省科学技術政策研究所「科学技術イノベーション政策のための科学」におけるデータ・情報基盤構築の推
進に関する検討」、NISTEP NOTE（政策のための科学） No.3（2012.11）
- [4] 黒沢努、他「JST ファンディング情報のデータベース化（JST-FMDB）とその活用法」、情報管理 Vol.58 No.4（2015）
- [5] 特許庁「特許ステータスレポート 2016」（2016）
- [6] 難波英嗣、他「学術論文の国際特許分類への自動分類」、Japio 2008 YEAR BOOK（2008）