企業技報の引用解析

～技報の引用情報から何が見えてくるのか～

○黒沢 努, 堀江 隆, 伊藤 祥, 住本 研一[1)], 伊藤 多一, 岩崎 哲, 西村 勇[2)]
李 星愛[3)]
科学技術振興機構[1)], 株式会社アイズファクトリー[2)], 株式会社オープンナレッジ[3)]
〒102-0081 東京都千代田区四番町五番地3
Tel: 03-5214-8402 FAX: 03-5214-8470
E-mail: tkurosaw@jst.go.jp

Citation analysis of the report of private companies:

What will come out from the citation of technical report?

○KUROSAWA Tsutomu,HORIE Takashi,ITO Sachi,SUMIMOTO Kenichi[1)],
ITOH Taichi,IWASAKI Tetsu,NISHIMURA Isami[2)], LEE Sungae [3)]
Japan Science and Technology Agency(JST) [1)], i's FACTORY co., ltd.[2)], Open Knowledge Corporation[3)]
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
Phone: +81-3-5214-8402 Fax: +81-3-5214-8470
E-mail: tkurosaw@jst.go.jp

【発表概要】

企業技報は、企業の CSR やブランド戦略、営業ツールとして刊行される一方で、各記事は特許出願後や学会発表後に関連記事が掲載される等、技術情報として、技術が製品化されていく流れを追うための重要な情報ソースである。

すなわち、企業技報はイノベーションの出口側における情報が掲載されていると考えられ、これらの引用情報を解析することで、企業が技術開発において、どのようなリソースを参照し、どのような機関とコミュニティを形成しているか、その関連性を解析することが可能と考えられる。

そこで、本調査では、企業技報の引用情報を約 10 年分(413 誌、約 35,000 記事、約 250, 000 引用情報)遡って入力し、出来る限り JST 所有の書誌情報との同定を試み、引用情報の解析を試みた。企業技報には、各業界の専門雑誌が引用されており、また、大学発の成果を企業が引用し、さらに大学や高専等で引用される等、産学の研究成果が循環型で融合されていることが分かった。

【キーワード】

企業技報, 引用解析、書誌同定、編集距離、知の構造化、可視化

1. はじめに

JSTはこれまで国内約12,000タイトル、海外約4,400タイトルの資料を収集し、毎年160万件規模の文献データベースを作成しており、現在、約3,000万件規模を蓄積しており、JDreamやJ-GLOBALといったWEBサービスを通じて、国内の研究者・技術を中心に提供されている。

近年は、データの増大に併せ、データを解析・可視化し、そこから知見を見出すことが主流となっており、海外のサービスも解析・可視化を検索に組み合わせたサービスが主力となり、その中でも引用解析が重要なファクターとなっている。

しかしながら、JSTの文献データには引用情報が搭載されておらず、国内文献の引用解析はJ-STAGEや国立情報学研究所のCiNiiに搭載されている引用情報のみが利用可能な状況である。

本報告では、今回初めて企業技報の引用情報を入力し、JST収録の書誌情報との同定作業とその解析を試み、同定作業を通じて分かった様々な問題点と引用解析から見えてくる結果について報告すると共に国内文献の重要性を考察したい。

2. 引用情報データの作成

本調査で対象とした資料は、2002年～2010年に発行された企業技報413誌、約35,000論文(引用情報数:約250,000件である。

これらのデータ入力にあたっては、JST情報資料館(成増)でOCRを中心に引用情報の作成作業を実施。約3か月間で引用情報を整備した。

3. 引用情報の書誌同定手順

作成された引用情報を解析するため、以下のステップで引用情報の書誌同定を実施した。

(1) 引用情報の分類
引用情報の全体像を把握するため、論文、特許、書籍、規格等、大まかに分類した。分類には、それぞれの情報ソース毎に引用パターンを表す検索集合を定義し、マッチングを実施。

(2) 同定手法の調査
(1)で分類した各種情報ソースを同定するための一致判定条件や類似度指標を設計。

(3) 同定作業の実施
(2)で調査した同定手法により、同定処理を実施。

(4) 実施結果の検証
サンプリング調査により同定精度の評価・検証を実施。

4. 引用情報の分類

分類結果は表1の通り。論文の占める割合が73%、会議録が11%、両者合わせて84%と大半を占める。

また、書籍は全体の6%程度で、その7割以上が国内書籍である.

特許は少なく、全体の1%程度で、その約8割が国内特許である.

企業技報はその性格上、当該技報が引用されるケースが多く、引用論文全体のうち、技報が占める割合は、約7%程度という結果となった。

表1 引用情報分類結果

▼引用文献リスト(全件)

データ種別	件数	構成比率	国内	海外	国内比率
論文	183,778	72.5%	100,969	82,809	54.9%
会議録	28,347	11.2%	16,333	12,014	57.6%
書籍	15,506	6.1%	11,353	4,153	73.2%
WEBサイト	13,045	5.1%	6,381	6,664	48.9%
規格	5,353	2.1%	5,353	0	100.0%
ハンドブック	3,323	1.3%	2,805	518	84.4%
特許	2,520	1.0%	2,021	499	80.2%
カタログ	1,224	0.5%	1,173	51	95.8%
新聞	266	0.1%	201	65	75.6%
一般雑誌	176	0.1%	171	5	97.2%
合計	253,538	100.0%	146,760	106,778	57.9%

うち、技報 12,141件(6.6%)

5. 書誌同定アルゴリズム

JSTでは、これまでJSTリンクセンター(現ジャパンリンクセンター(JaLC))における引用・被引用リンク、ReaD研究者

情報の成果論文とJ-GLOBAL登載の書誌データとの同定等、異なる媒体に登載された書誌情報間の同定を実施してきた。これらは、アプリケーション毎に改善が施されていたため、今回は、これまでのアルゴリズムを踏まえつつ、企業技報に適した改善を行った。

(1) J-GLOBALの書誌同定[1)]
- ○複数項目のマッチングから総合的かつ機械的にマッチング判定.
- ○タイトルの一致には、n-gramによる段階的判定を採用(50%以上一致、90%以上一致).
- ○False Negative(本来結びつけるべきもの同士を結び付けない誤り)は少ない反面、False Positive(本来結び付けてはいけないもの同士を結びつける誤り)も多い。

(2) 今回の書誌同定
- ○引用情報にはタイトルが含まれないケースが多いため、引用情報の書誌同定は、JSTリンクセンターの書誌同定に近い。
- ○JSTリンクセンターのアルゴリズムを参考にしつつ、False Positiveが少なく、事前に分類した引用情報毎にチューニングしたアルゴリズムで処理することで、False Negativeを抑える方向で実施。

6. 同定結果

5. の方針のもと、書誌同定アルゴリズムを開発し、試行錯誤をしつつ、書誌の同定作業を行った。

企業技報の引用情報は、予想以上に表記揺れや表記パターンの異なりが大きく、特に、書誌情報を一意に特定するためには、資料名、巻、号、ページを特定することが基本的な要素となるが、資料名の略記やその他の特定要素の欠落、記述間違いが大きく、その結果として精度が不充分な結果となった。

(1) 和文誌
- ○約3割が同定済(36.5%)

(2) 英文誌
- ○約3割が同定済(36.7%)

(3) 特許
- ○約5割が同定済み(45%)
  - ※同定できないものの多くは収録対象外、海外特許等。

(4) 書籍
- ○約1割の同定に留まる
  - ※著者名や文字の区切り判断が困難なことが主要因。書籍は同定するための材料が少なく、区切りが不規則なため、文章全体で類似度を判定する方法が有効か。

7. 同定精度向上に向けた改善方策

同定精度向上に向けて対策としては、以下が有効であると考えられる。

(1) 資料略記辞書の整備

例) 資料略記の表記揺れ

機論 → 日本機械学会論文集

信学技報 → 電子情報通信学会 技術研究報告

回塑加連講論→ 塑性加工連合講演会講演論文集

(2) 引用情報の表記ルール(パターン)の蓄積と同定アルゴリズムへの反映

○ 引用情報の要素不足

「年」「巻」「号」「ページ」の情報無しでは情報が少なく検出できない

例)「豊倉・金元・八田, 遠心ターボ機械の戻り流路川円形翼列に関する研究, 機論」

○ 引用情報の間違い

引用情報の「巻」「号」「ページ」いずれかが違うため、正しい書誌情報を絞り込めていない

○ アルゴリズムの改善

記事タイトル(n-gram)の類似度を含めた総合アルゴリズムの検討。

※ 処理時間の勘案が必要。

8. 引用情報の解析

書誌同定によって得られた引用情報67,302件(和文誌:36,872件、英文誌:30,430件)の解析により、新たな知見を見出すべく、解析・可視化を実施。

ソフトはOpenKnowledgeViewer(株式会社オープンナレッジ)[2]を使用した。

(1) 技報に含まれる引用情報の傾向

概況は以下、図2, 3の通り。

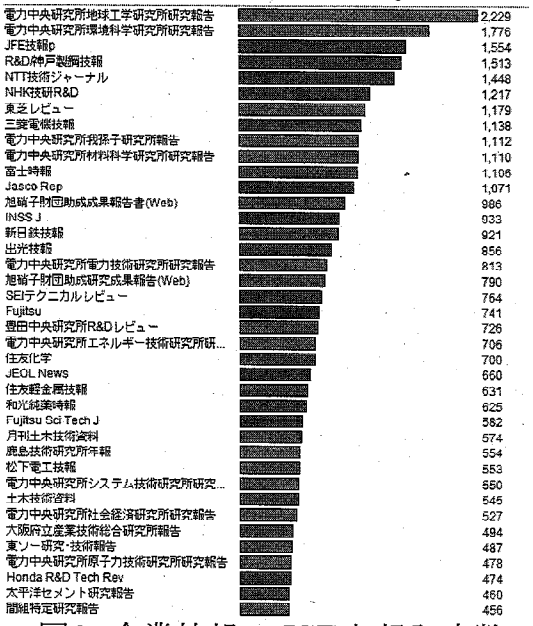


図2 企業技報のJST収録記事数

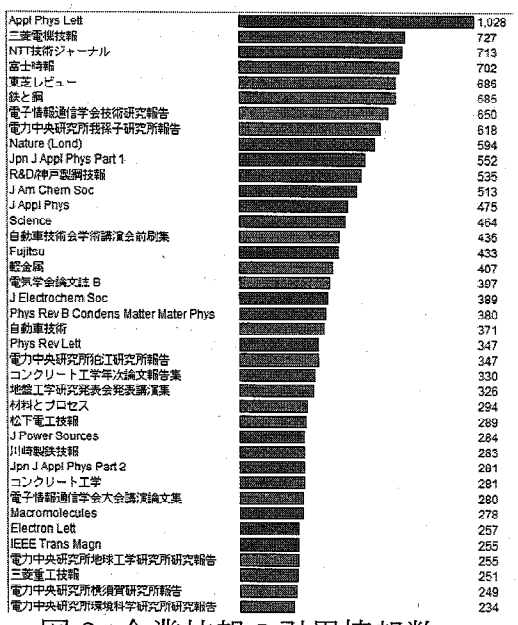


図3 企業技報の引用情報数
(本報告のために整備した引用文献)

企業技報に引用される資料のうち、上位100位以上の資料のJ-STAGE搭載率は36%である。企業技報から高頻度で引用される資料は産業界のビジビリティが高いと推定され、今後の収録基準に重要な視点である。

(2) 大学の研究業績の引用傾向

技報が引用する各資料について、引用先が大学である文献の割合を図4に示す。外国誌を引用する場合は、大学の先生の業績を引用する傾向が強い。

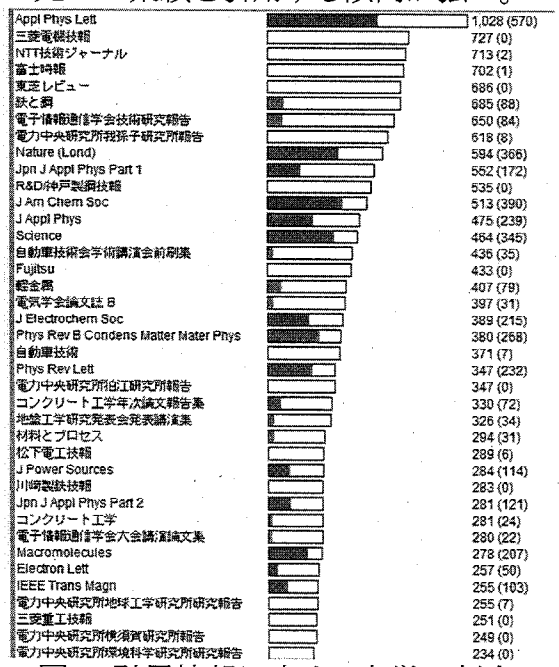


図4 引用情報に占める大学の割合
(塗りつぶし範囲が大学の共著)

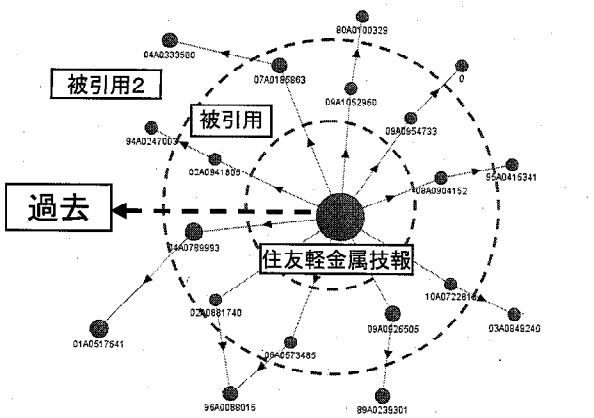
(3) ネットワーク解析

図5は住友軽金属技報のある記事を中心に引用関係を示したものである。

「金属」や「軽金属」という専門雑誌が産学のコミュニケーションの場となっていることが推定される。

また、所属機関別に見るとも元々は大学の成果(被引用2)であったものを高専、社団法人、企業等が引用(被引用)し、更に、住友軽金属技報で企業が引用するというような、産学の研究成果が循環

型で融合していることが分かった。(中央が注目している文献。同心円の外側に行くほど過去に遡ることを示す)



ab 【被引用】資料名		ab 【被引用2】資料名	
軽金属	14 (100%)	軽金属	9 (64.3%)
		住友軽金属技報	2 (14.3%)
		金属	2 (14.3%)
		(データなし)	1 (7.1%)

ab 【被引用】機関名		ab 【被引用2】機関名	
住友軽金属工業 研開セ	8 (40%)	住友軽金属工業 技研	5 (27.8%)
住友軽金属工業(株)研究開発センター	4 (20%)	東京工大 大学院	3 (16.7%)
旭川工高専	2 (10%)	東京工大 大学院理工学研究科	3 (16.7%)
東京都産業技術専	2 (10%)	東京都工高専	2 (11.1%)
社団法人日本アルミニウム協会	2 (10%)	(データなし)	1 (5.6%)
住友軽金属工業株式会社研究開発センター	1 ...	住友軽金属工業 研開セ	1 (5.6%)
名古屋大 大学院工学研究科	1 (5%)	大阪府大 工	1 (5.6%)
		東大大学院	1 (5.6%)
		東大工	1 (5.6%)

図5 引用情報のネットワーク解析

10. 結果と考察

(1) 企業技報の引用情報の種類

90%近くが文献由来(論文、会議録)特許やカタログ等は少ない。

(2) 企業技報の引用同定

現時点で約 36%が同定可能(論文)。同定できない要因は資料名の揺れ(略記等)や不備によるものが大きい。今後、略記辞書が整備できれば約 50%迄改善可能と推定)。

それ以外に発行年、ページ等、情報を一意に特定する上での重要事項の漏れ(全体の約40%と推定。タイトルがあるものはタイトル類似ロジックにより改善と推定)

但し、これらは対処療法であり、根本から解決するためには、技報等の投稿の際、JST の書誌を API 等を通じて受け取り、ID と共に完全な書誌とすることが望ましい。

今後、これらの結果については、過去の同定アルゴリズムと比較検証した上で、JaLC の WG に報告し、今後の改善に役立てることを予定している。

最終的には JaLC 経由で投稿時から、全ての文献に DOI が付与され、一意性が保証されることが望ましい。

(3) 引用解析から分かること

引用解析を通じて、企業の研究者・技術者に近いと考えられる資料が分かり、学から産へのイノベーション創出を目指す JST として、電子化を優先的に進める際の手がかりとなった。

また、同じ業界でも他社からの引用が多いもの、少ないものがあり、競合からの注目度等も解析可能である。

今後、更なる解析を進めることにより、各機関(JST 等)の研究成果の波及効果等についても深堀りしたい。

11. おわりに

技術の進歩によって取り扱うことができるデータの範囲が広がり、これまでのデータに加え、リアルタイム性の高いデータ(ビッグデータ)を組み合わせて最適解を求める動きが活発化してきた。

これら情報に基づく意思決定のためには、判断基準となる信頼性の高い情報が不可欠であり、国内文献の引用情報についても、今が整備のときではないかと考える。

12. 参考文献

[1] 松邑勝治,黒沢努,関根基樹,矢口学,植松利晃,加藤治,「J-GLOBAL」試行版(β版)の構築と今後の展望,情報管理,Vol. 52,(2009),No.3,P150-157

[2] OpenKnowledgeViewer (http://openknow.com/)