

文献データベースのAPIを活用した 独自の論文データベースの構築と研究成果の可視化

○野々村 美貴^{A)}、甲斐 誉人^{B)}、高山 有道^{C)}、坂本 隆一^{C)}

自然科学研究機構 核融合科学研究所 技術部^{A)}

WDB 工学株式会社^{B)}

自然科学研究機構 核融合科学研究所 研究部^{C)}

1. はじめに

核融合科学研究所（NIFS）では、論文情報システム NIFS Article Information System（NAIS）^[1]という研究業績データベースを内製・運用している。NAISでは、論文情報の登録を研究者に委ねているため、登録状況にばらつきがある。このため、NAIS 単独ではすべての研究成果を網羅することができない。そこで外部文献データベース（Scopus^[2]、Web of Science^[3]、ORCID^[4]）のAPI^{*1)}を利用して、論文情報を集約しNAISと統合することで、より多くの研究成果を可視化できるようにした。本稿では、論文情報の自動更新・検索機能を備えた独自の論文データベース NIFS Research Activity Database（NRAD）の構築とその活用方法について報告する。

*1)Application Programming Interface の略称で、異なるアプリケーションの間でデータやプログラムの機能を共有できるようにしてくれる仕組み

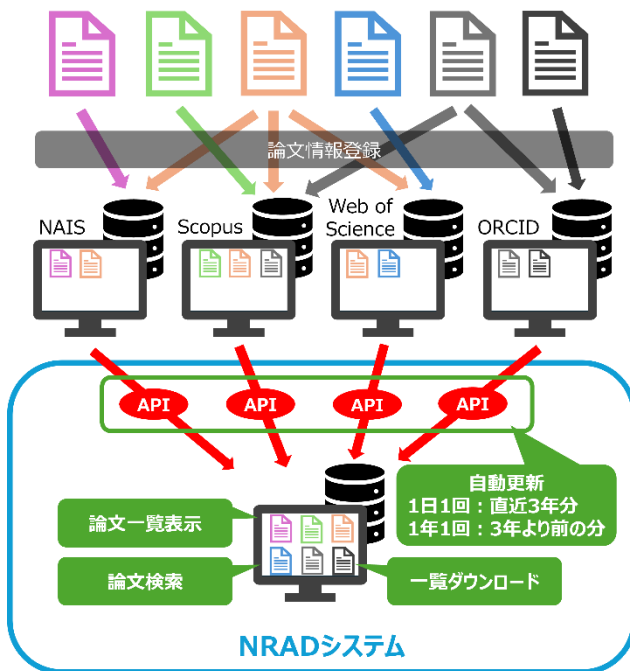


図1 NRADシステム全体概要

2. NRAD システムの開発

外部文献データベースから論文情報を自動的に登録・更新し、研究者ごとに論文情報を可視化・検索する機能を有するシステムを構築した。本システムの全体概要を図1に示す。

2.1 文献データベースからの論文情報取得

文献データベースは、それぞれ独自の方法で構築されている。現状ではすべての論文情報を網羅した文献データベースは存在しないため、複数の文献データベースから論文情報を収集することとした。表1に示す理由により文献データベースを選定した。

表1 文献データベースの選定理由

DB名	選定理由
Scopus	多くの文献が登録されている 詳細な情報を取得できる
Web of Science	Scopus で登録されていない文献が登録されている
ORCID	研究者が独自で文献を登録している
NAIS	研究者が独自で文献を登録している 会議論文情報が登録されている

論文情報は、各文献データベースのAPIを利用して取得している。APIのしくみを図2に示す。

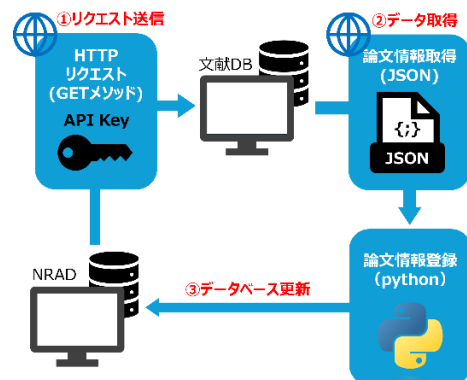


図2 APIのしくみ

表2 各文献データベースのAPI

DB名	API名	リクエスト使用制限	クエリパラメータ	応答フォーマット
Scopus	Scopus Search API	20,000件 / 週 9件 / 秒	所属ID 個人ID 出版年 著者ID等	JSON
	Abstract Retrieval API	10,000件 / 週 9件 / 秒		
	Author Retrieval API	5,000件 / 週 3件 / 秒		
	Affiliation Retrieval API	5,000件 / 週 9件 / 秒		
Web of Science	Web of Science Starter API	5件 / 秒 1,000件 / 日 50文献 / 件	所属 出版年 個人ID等	JSON
ORCID	Public API	12件 / 秒	個人ID	JSON

APIにリクエストを送信する際に指定する項目は次の通りである。

1. ヘッダー
 - 応答フォーマット
2. クエリパラメータ
 - API Key (事前に取得する必要あり)
 - 個人ID または 所属ID
 - 出版年 等

表2に示す通り Scopus と Web of Science は、一定期間内での使用制限があるため、1日に1回行う自動更新は直近3年分とした。それ以前の論文は新規登録や更新があまりないため、1年に1回自動更新を行うこととした。

2.2 論文情報の登録・更新

APIで取得した論文情報は、python スクリプトでデータベースに登録・更新を行っている。このスクリプトは処理負荷が高く、実行に時間がかかるため、他に影響の少ない夜間の時間帯に実行している。スクリプトを実行する際に注意する点は次の通りである。

- APIの使用制限を超えないように、リクエストごとに間隔を空ける
- スクリプト実行時に CPU 使用率が 100%に達しないように、cpulimit コマンドを用いて制限している (現状は 60%程度に抑えている)

毎日の自動更新にかかる時間は日によって違うが、ある日の例を表3に示す。

表3 自動更新の所要時間と件数

DB名	所要時間	件数 (更新数 / 論文数)
Scopus	約 30 分	2023年 0 / 228 件 2024年 1 / 286 件 2025年 1 / 52 件
Web of Science	約 4 分	2023年 0 / 167 件 2024年 1 / 259 件 2025年 1 / 33 件
ORCID	約 4 分	2023年 0 件 2024年 0 件 2025年 1 件
NAIS	約 11 分	1968~2023年 0 / 16241 件 2024年 4 / 285 件 2025年 3 / 34 件

2.3 論文情報の同一判定

複数の文献データベースから論文情報を収集するため、同じ論文情報が登録されることがある。このため、論文情報を比較しデータを統合する必要がある。NRAD システムでは、次の条件が揃えば同一論文と判定することとした。

- DOI^{*2)}が一致
- タイトルの 90%以上が一致
- ジャーナル名が一致

^{*2)}Digital Object Identifier の略称で、インターネット上にある電子化されたコンテンツに恒久的に付与される国際的な識別子

この方法だけでは論文情報の統合が不十分であるため、表4に示す通り追加の対応が必要となった。

表 4 システムでの自動同一判定の問題点

問題点	対応内容	対応件数
一つの論文に対して、複数の DOI が存在しており、その内の 1 つにリダイレクトされている	リダイレクトされている DOI の論文を同一判定から除外	8 件
一つの論文に対して、Paper と Proceedings がある	Proceedings を除外	15 件
一つの論文に対して、Erratum, Editorial 等があり複数の DOI が存在している	Erratum, Editorial 等の論文を除外	69 件
DOI が間違っている	各文献データベースに修正依頼	39 件
タイトルは同一だが別論文である	内容を確認し、別論文とする	14 件
同姓同名の人物の論文が混在している、プロフィールが複数ある	同一人物か、プロフィールの確認をする	61 名

研究者数：119 名、登録年数：58 年（1968～2025） 総論文登録数：31,300 件、集約後論文数：12,200 件
（2025/2/20 現在）

2.4 研究成果の可視化

NRAD システムで統合した論文情報を、研究者ごとに Web ブラウザ上で確認できるようにした。各研究者の Web ページでは、論文情報を 4 つのカテゴリ（Refereed Papers、Presentation、Book、Others）に分類し表示した。各カテゴリでは、検索・ソート・ダウンロード・ページ切り替え機能が利用できる。これらの機能は、JavaScript ライブラリ jQuery^[5]のプラグイン tableorter^[6]で実装した。図 3 に各研究者の Web ページイメージを示す。



図 3 各研究者の Web ページイメージ

3 NRAD システムの活用

NIFS では、NRAD システムで統合した論文情報を次の用途に活用している。

- 研究成果の公開
（研究者毎の論文情報を JSON 形式で研究者紹介ページに送信）
- 研究成果の報告等に活用
（研究者毎の Web ページより論文情報一覧をダウンロード）
- 研究者の業績評価に利用

4 まとめ

NAIS では、論文情報の登録を研究者に委ねていたため登録状況にばらつきがあったが、NRAD システムでは複数の文献データベースから論文情報を収集することで、より多くの研究成果を網羅できるようになった。これにより研究者の登録作業の負担を軽減することができた。また、研究成果を統合し可視化できるようになったことで、論文情報の公開や業績評価が公平に行えるようになった。

今後は、システム運用者の要望を受け、論文情報の検索機能の強化、分類作業の機能追加を行う予定である。

参考文献

- [1] NIFS Article Information System
<https://nais.nifs.ac.jp/article/center>
- [2] ELSEVIER Scopus
<https://www.elsevier.com/ja-jp/products/scopus>
- [3] Clarivate Web of Science
<https://www.webofscience.com/wos/>
- [4] ORCID
<https://orcid.org/>
- [5] jQuery
<https://jquery.com/>
- [6] tableorter
<https://mottie.github.io/tableorter/docs/>