

レポート

科学技術白書検索システムの紹介

第2研究グループ 客員研究官 岸本 晃彦、総括主任研究官 富澤 宏之

概要

科学技術・学術政策研究所（NISTEP）で構築・開発した「科学技術白書検索システム」の利用方法を紹介する。活用事例として、理工系人材に関する政策史を研究する場合、及び、コロナウイルスを含む感染症についての政策動向を調査する場合をとりあげて具体的な利用方法を示すとともに、政策研究への活用可能性と、データ・情報基盤の今後の方向性について述べる。

キーワード：政策立案，科学技術白書，政策史，コロナウイルス

1. 科学技術白書検索システム：背景と概要

昭和33年以降発行され、現在は文部科学省が取りまとめている科学技術白書（以下、「白書」という。）は、日本の科学技術の動向を歴史的に俯瞰して把握することができる貴重な資料である。科学技術・学術政策研究所（NISTEP）では、科学技術政策研究を始め様々なニーズに活用できるよう「科学技術白書検索システム」を構築し、公開した¹⁾。

本システムの特長のひとつは「あいまい検索」と「部分一致検索」の両者を実装していることである。「あいまい検索」は関連する語句を検索するので、正確な検索語が思い浮かばないときでも、それに近い語句を入れて一定数（100件）検索でき、考察することができる。そこから探したい語句が見つければ、「部分一致検索」でその語句の使用時期や頻度などの統計的なデータも得られる。なお、この「あいまい検索」の機能は、白書の全文や研究機関のwebサイト等の研究概要やプレスリリースなどのテキストデータを用いて、様々な語句が同時に出現する頻度を測定し、それに基づいて、検索語に“近い”（あるいは“関連性が高い”）語句をリストアップすることにより実現している。

図表1に白書検索システムのトップページを示した。「収録白書一覧」は収録した白書の一覧である。また、「白書検索を利用する」をクリックすると「白書検索」の画面に移る（図表2の中央）。さらに、「白

書詳細」により白書の全内容を閲覧することや、「段落抽出」ができる。これらが白書の内容を「検索・閲覧」できるページである。

本システムには、白書の内容をグラフ等で表示する「分析ツール」がある。すなわち、①語句の出現頻度などから算出した重要性を概観する「キーワードマップ」、②注目する語句を選び、その語句の出現頻度の変化をグラフ化する「フリーワード出現回数分析」、③文書のつながりを年版順に追っていく「関連文書時系列分析」といった分析ツールを実装している。以下、具体的な利用例を挙げて説明する。

2. 事例1：科学技術政策の歴史的研究への活用

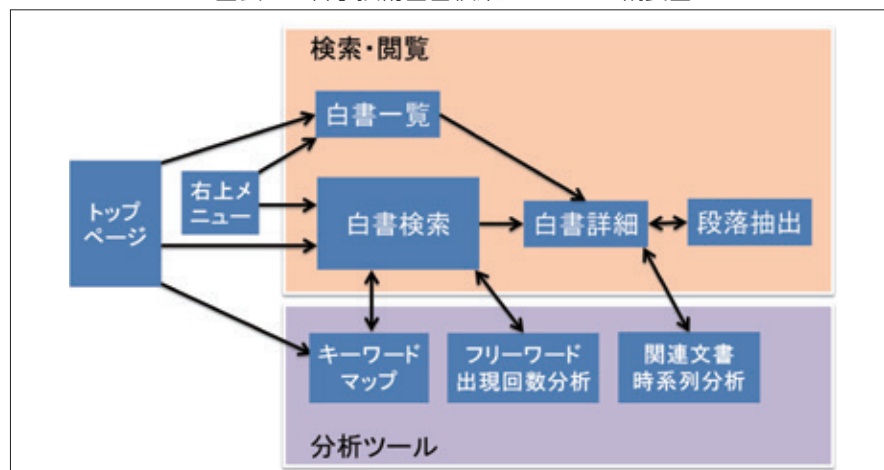
日本の科学技術イノベーション政策では、人材育成に関して、産業や社会の変化に対応した人材を供給するシステムとなっているのか、などが論点になることが多い。このような問題に関しては、過去に遡って、現在のシステムがどのように形成されたのかを考察することが有用な場合がある。そこで、高度経済成長期における長期計画として有名な「国民所得倍增計画」（1960年12月27日閣議決定）のもとで、科学技術人材育成システムが形成された状況を調べてみる。

トップページから白書検索のページに入り、「所得倍增計画」というフリーワードの部分一致検索を行い検索結果の一部を図表3に示した。図中の右の方に

図表 1 科学技術白書検索のトップページ



図表 2 科学技術白書検索システムの概要図



図表 3 “所得倍増計画”の検索結果画面（一部のみ）



「18 件中 1 - 18 件目を表示」とあるように、この検索で 18 件ヒットしている。それを「古い順」に表示させると、「所得倍増計画」の語が出現する最も古い白書は昭和 37 年版（1962）であることが分かる。同年版では 6 件の出現箇所があり、当時、「所得倍増

計画」に呼応して実施された様々な科学技術政策の内容を知ることができる。

図表 3 で、最初の検索結果として表示されているのが「文部省の理工系学生増員計画」についての節であり、これが今回、調べている科学技術人材に関する

記述である。青くハイライトで表示された「所得倍増計画」の前後それぞれ100字が示されている。これを選択すれば詳細画面を開くことができる(図表4)。図表4の「内容」の部分に、当該の白書の本文が表示されており、オリジナルの文章や図表を見ることができる。また、その上方にある「パラグラフ情報」は、当該箇所の全体の中での位置付けを知るために有用である。すなわち、昭和37年版(1962)の白書では、総論第Ⅱ部「科学技術発展の基盤」の第2章のテーマが「人材養成」であること、また、その中の「Ⅱ.」という節で「科学技術者の養成計画」がとりあげられていることが分かる。この「パラグラフ情報」は、検索でヒットした部分だけでなく、それを含む節や章の全体を読むためのガイドの役割も持っている。すなわち、検索でヒットしたパラグラフの上位レベルの「Ⅱ.」をクリックすることにより、「科学技術者の養成計画」を扱った節の冒頭に簡単にアクセスできる。

以上では、説明を簡単にするために、ひとつの検索語により、求める情報が簡単に得られた例を示したが、実際には、どのようなキーワードで検索すれば良いのか、当初は曖昧である場合も多い。しかし、この例の場合、部分一致検索で「所得倍増」と「人材」のAND検索、あるいは「理工学系学生増員計画」の語を知らなくても、あいまい検索で「理工学系増員」や「理工学系倍増」を検索すれば、同様の結果が得られる。

3. 事例2：コロナウイルスと感染症に関する分析例

2019年末に中国武漢市で発生した「新型コロナウイルス感染症」は、日本のみならず世界中を巻き込み

猛威を振るっている。連日、新型コロナウイルス関連がトップニュースで報じられ、多くの国で都市封鎖された。日本でも緊急事態宣言が出されるなど、人の移動が大きく制限され、経済的な打撃も計り知れない。

そこで、これまでのコロナウイルスや感染症に関する科学技術政策の取組について白書での検索・分析を試みた。

「白書検索」のページで「コロナウイルス」を部分一致検索で検索すると、2007年版に1回、2015年版に8回使われていた。2015年版の「SARS」の記事は、2007年版の内容も含んだ形で詳しく記載されている。

この2015年版の「SARS」の記述によれば、① SARSは2002年11月16日に患者が初めて確認され、② 2003年7月5日に、世界保健機関(WHO)によって終息宣言が出されるまで、③ 29の国と地域で、8,096人の感染者と、774人の死亡が報告された。④ 2003年3月、WHOはこの原因不明の重症呼吸器疾患を、Severe Acute Respiratory Syndrome(SARS)と命名して研究ネットワークを創設し、約1か月という短期間で、新型のコロナウイルスがSARSの病原体であることが突き止められた。⑤ 国立感染症研究所等と栄研化学株式会社は、1時間以内に検出できる迅速簡易診断キットを開発し、水際防止に役立てられた。⑥ 理化学研究所はSARSコロナウイルスの増殖阻害の可能性のある物質を発見、シミュレーションで候補化合物を選択、その後治療薬の実用化が進められている。⑦ 幸いにも、我が国ではSARS患者は発生しなかったが、SARSの流行は、未知のウイルスによる世界的な感染爆発への緊急危機対応における、病原体の特定と、診断・治療・予防方法の開発のための研究開発の重要性を浮き彫りにした、と結論付けている。

図表4 「文部省の理工系学生増員計画」の詳細画面(一部のみ)

白書詳細

段落抽出

▼

分析ツール

▼

文科省ウェブサイトで見える

▼

パラグラフ情報

昭和37年版(1962) 科学技術白書

総論

第Ⅱ部 科学技術発展の基盤

第2章 人材養成

Ⅱ. 科学技術者の養成計画

1. 文部省の理工系学生増員計画

内容

第Ⅱ部 1. 文部省の理工系学生増員計画

このような事態に対応するために、文部省では、昭和36年2月に理工系学生定員増加計画を策定した。本案は理工系学生を昭和36年度から44年度の間に1万6,000人増員しようというものであった。(4年制大学については昭和42年までの7年間で、短期大学については昭和44年度までの9年間に増員を達成する。)この案に対し、科学技術庁長官はさらに初年度における大巾な増員計画を検討する必要がある旨の勧告を行なった。

「新型コロナウイルス」は2019年末に出現したので、令和元（2019）年版までを検索対象とする現時点（2020年7月）の「白書検索」には入っていない。しかし、2020年6月、令和2（2020）年版の白書が文部科学省から公開された²⁾。なお、このサイトからは、NISTEPの本「科学技術白書の検索システム」へのリンクが張られている。この文部科学省公開の白書最新版で、「コロナウイルス」、「新型コロナウイルス」、「新型コロナウイルス感染症」及び「感染症」の4つの語句が出現した回数を調べたところ、それぞれ、64回、60回、36回、97回であった。「コロナウイルス」について2020年版（64回）は2015年版（8回）の8倍と激増しており、新型の「コロナウイルス」への注目度がいかに高いかが分かる。

次に「白書検索」の「キーワード出現回数分析」を用いて「感染症」の出現頻度を調査した。「白書検索」では、「施策編」と「施策編以外」とを分けて表示している。「施策編」は、科学技術基本計画に沿って施策が実施されていることを確認するもので、最近の白書では第2部に掲載され、年版を通じて継続的な内容が多くなっている。一方、「施策編以外」は、その年の話題を特集したものなどで、最近の白書では第1部に掲載され、年版によってそれぞれ特色がある。「感染症」について2019年版までの「白書検索」の結果に、上記令和2（2020）年版の白書の結果を加えて図表5に表示した。

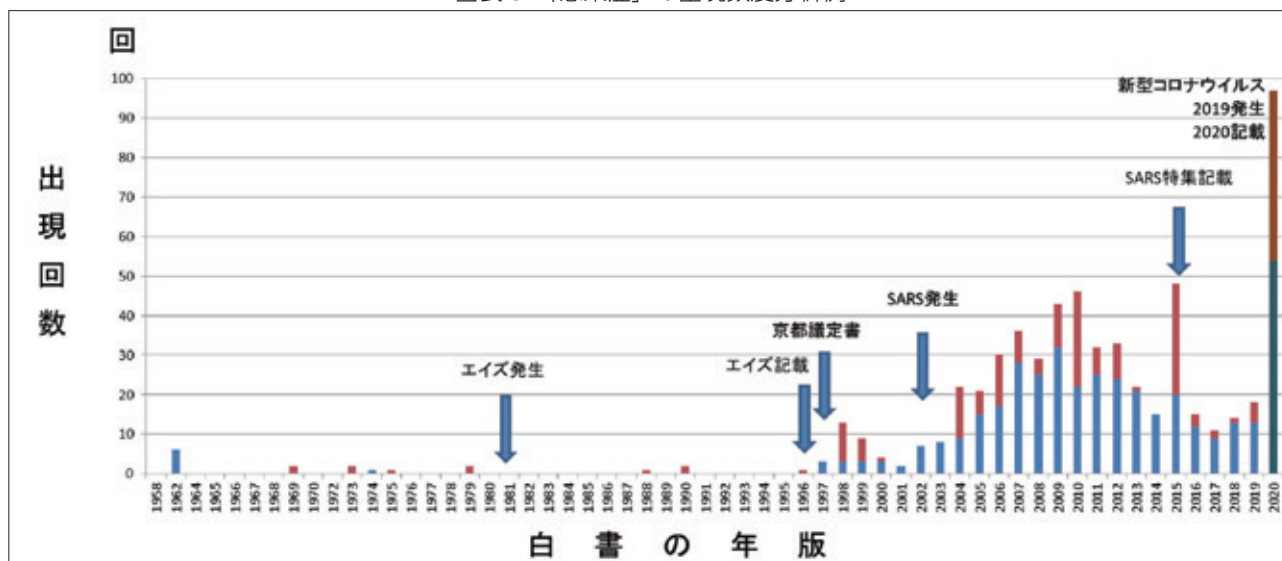
「白書検索」による「感染症」の検索結果には、「感染症」の語句が出現した前後それぞれ100文字が記載されているのでヒットした箇所の概略が分かる。ま

た、白書の年版、編、章、節、及びその下位の記事（いわゆる「パンくずリスト」）も記載されるので、白書における位置付けも把握できる。

これを使うと①「感染症」が最初に出現するのは白書の刊行を開始した少し後の1962年版からで「ウイルス感染症」といった記事の中で使われている。②1996年版白書では「エイズ」の記事が「感染症」関連としてヒットしている。③1997年12月、「京都議定書」が採択され、白書では地球規模の対応策として、「地球温暖化」と合わせて「感染症」も議論されてきた。④2002年11月の「SARS」の発生から他の感染症も併せて扱う「新興・再興感染症」の記述が多くなり、⑤「新興・再興感染症拠点形成プログラム」が2007～2011年版まで継続的に掲載されるなど、徐々に「感染症」の出現頻度が増加した。⑥「感染症」の出現頻度は2010年版をピークに徐々に少なくなったが、⑦2015年版には「施策編以外」に「SARS」の詳しい記述があるため、「新型コロナウイルス」が記載される以前では最も多い48回となった。⑧「新型コロナウイルス」が記載されている2020年版は94回で、2015年版に比べて倍増している。

2020年版の白書で「感染症」が多く使われている箇所は第1部（施策編以外）では「はじめに」（13回）、「新型コロナ感染症に関する研究開発について」（24回）、第2部（施策編）では「第3章 経済・社会的課題への対応」（46回）である。第3章の中の健康長寿社会の関係で、以前から掲載されている「ク 新興・再興感染症に関する研究」（11行）に新たに「ケ 新型コロナウイルス感染症への対応」（2ページ

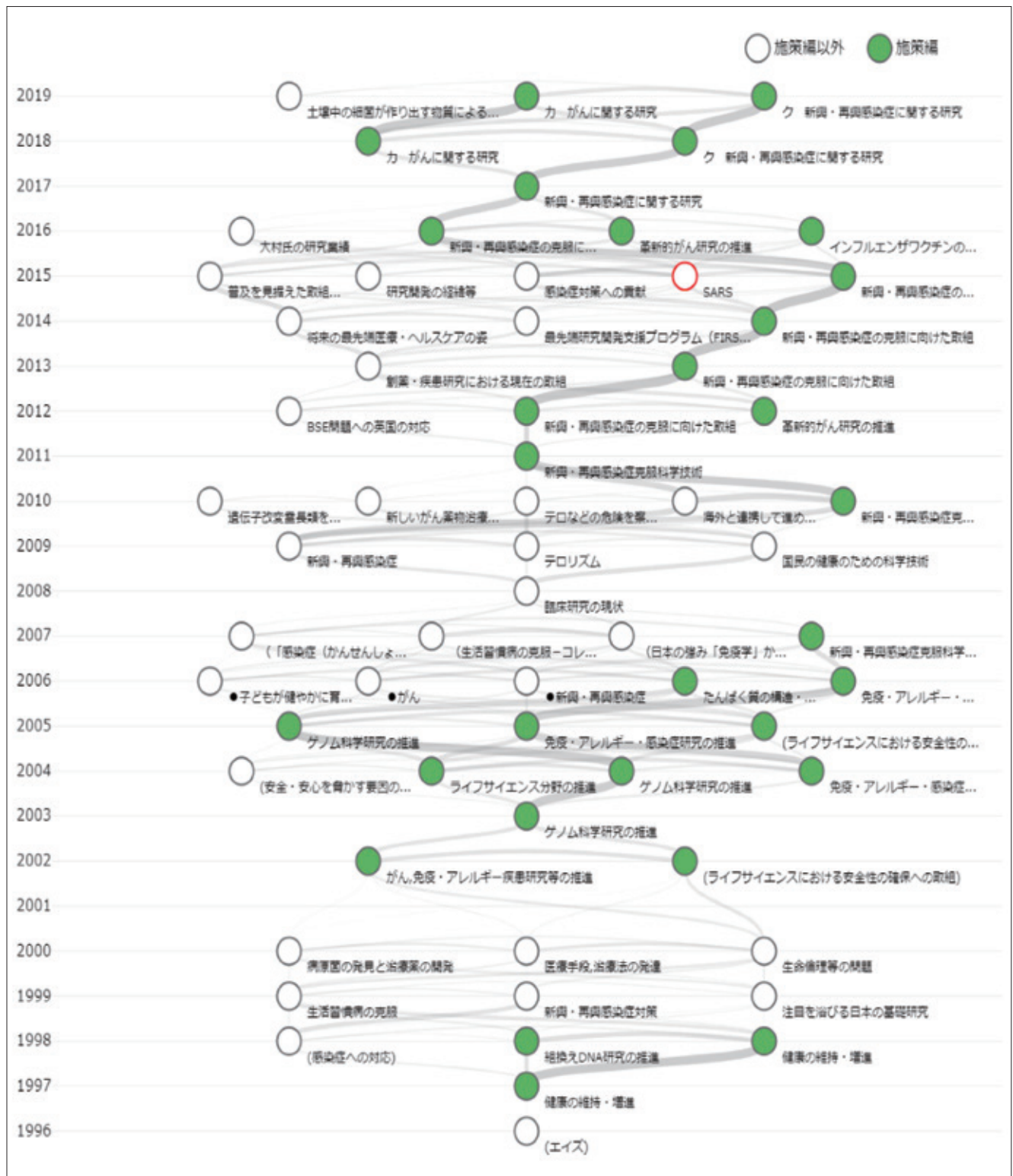
図表5 「感染症」の出現頻度分析例



青色：施策編、オレンジ色：施策編以外。2019年版までは「白書検索」の「キーワード出現回数分析」によるもので「感染症」出現回数分布検索 HIT 件数：229 件、キーワード総出現回数：499 回であった。

濃い色の2020年版は文部科学省のPDF版白書から手作業で算出。

図表7 SARS の 2015 年版の記事に関係の深い記事の最下層記事の時系列分析



参考文献・資料

- 1) NISTEP 科学技術白書検索 <https://www.nistep.go.jp/research-scisip-whitepaper-search>
- 2) 文部科学省科学技術白書 https://www.mext.go.jp/b_menu/hakusho/html/kagaku.htm
- 3) 統合イノベーション戦略 2020 <https://www8.cao.go.jp/cstp/tougosenryaku/index.html>