

PATENTSCOPE の化合物検索の検証 ～キーワード検索との比較

○石川 彰¹⁾

中央光学出版(株)¹⁾

〒105-0003 東京都港区西新橋3-11-1 建装ビルディング 4F

Tel: 03-6721-5561 FAX: 03-3436-2681

E-mail: akira.ishikawa@cks.co.jp

Evaluation of Chemical Compounds Search of PATENTSCOPE --- Comparison with Keyword Search

ISHIKAWA Akira¹⁾

Chuo Kogaku Shuppan Co.,LTD.¹⁾, Kenso Bldg.4F,

3-11-1, Nishi-Shinbashi, Minato-ku, Tokyo, 105-0003, Japan

Phone: +81-3-6721-5561 Fax: +81-3-3436-2681

E-mail: akira.ishikawa@cks.co.jp

【概要】

化合物の構造式の作図入力で検索ができることは、科学文献調査において有意義である。従来、専任のアナリスト(人)の力により構築されたデータが一般的であった。一方、最近、特許のデータベースでは人の力を使わず、機械的に検索キーを設定する方式のデータベースが上市されている。

本報告では、機械的に検索キーを設定した特許データベースのうち、世界知的所有権機関(WIPO)が提供するデータベース PATENTSCOPE について検索検証を行った。検索対象を農薬 amidosulfuron として、化合物検索とキーワード検索の結果比較を行った。

i) 化合物検索でヒットした案件は、99%超がキーワード検索でもヒットした

ii) キーワード検索でヒットした案件のうち、化合物検索でヒットしないものが約15%あった

キーワード検索の結果を精査したところ、化合物検索でもヒットすべきことが分かったが、上記 i) は妥当であり、一方、上記 ii) は好ましくない。ii) の原因は、化合物検索の検索キーを抽出する元になる公報データが、キーワード検索で検索するデータと異なるためであると推定する。この点が改善されることを期待する。

【キーワード】化合物検索, 検索キー, Chemical compound Search, Patent Search, Inchi key, PATENTSCOPE, Orbit-Intelligence

1. はじめに

化合物の特許検索データベースでは、化学情報協会が提供するCAplusが著名である。このデータベースは、化学構造式の入力により検索できる特徴がある。分類専任者「アナリスト」が構築したデータであり、信頼性が認められている^{1, 2)}。しかし、処理データ量が増加しつつある現状では、データベースを維持・拡張するには時間・労力に課題があると思われる。更にデータベース利用料が高価なため、多くの利用者が使いやすいとはいえない。

一方、世界知的所有権機関(以下、WIPOと略記)の提供するデータベースPATENTSCOPEは、2016年から化合物検索の機能搭載している^{3, 4)}。PATENTSCOPEでは、データ構築が、専任者のマニュアル作業を伴わず、機械力により構築される点異なる。図面が検索対象にならない、マークッシュ形式の記載にも対応できないなど、課題はあるが、期待度は高い。PATENTSCOPEにおける化合物検索機能の搭載以降、商用データベースでも、同じ方式の化合物検索の機能が搭載され始め、注目されている。データベースの比較検討が、各特許検索の研究会・協議会で行われている^{5, 6)}。

本報告は、これらの議論と同様にPATENTSCOPE方式の化合物検索に期待する立場から検証を行ったものがある。検索対象は、農薬成分の一つであるamidoflufenuron(除草剤の成分)で行った。

なお、検証は2018年11月時点で行ったものである。その後PATENTSCOPEはバージョンアップされたという情報があったが⁷⁾、本報告で見つかった課題は12月10日時点でも変わりが無い。

また、PATENTSCOPE以外のデータベ

ースの化合物検索においても、共通の課題が見つかっている。

2. 本化合物検索の原理

2.1 使用した特許データベース

本検討では、WIPO提供のPATENTSCOPE及び、確認のため、Questel社提供のOrbit-Intelligence(以下、Orbitと略記)も適宜利用した⁸⁾。

PATENTSCOPEの化合物検索の機能を表1に示す。

表1 PATENTSCOPEの主な仕様

PATENTSCOPE特徴	
サービス提供	世界知的所有権機関(WIPO)
利用可能言語(インターフェイス)	英語
対象公報	WO:1978- (英語、独語) ※ US:1979-
データダウンロード	指定項目のみ可能(上限10000件)
利用料金	無料(登録が必要)

※特許協力条約出願(WIPO出願)

2.2 PATENTSCOPEでの検索

PATENTSCOPEの利用について、注意事項が分かりやすく報告されており⁹⁾、これを参考にした。5つの検索モードのうち構造化検索(Structural Search)は入力方法が平易で使いやすいが、履歴検索の機能がない。一方、詳細検索(Advanced Search)ではコマンド検索が可能であり、しかも長文の検索式を入力できる。本報告では詳細検索を使用した⁷⁾が、1ショットの検索式の中に、履歴演算の内容を入れた長めの検索式を作成して行った。

また、検索のヒットリストでは、クリックすると、案件毎に、書誌、明細書、請求項、検索キーとして抽出された化合物等を表

示する機能がある(図1(a)(b))。この機能を使い解析を行った。

(a) 公報閲覧用の表示

(b) 化合物閲覧用の表示

図1 PATENTSCOPE のヒット案件の表示

3. 結果と考察

3.1 原理

PATENTSCOPE 方式の化合物検索の原理を図2に示す。各公報に対しては、公報のテキスト中から化合物の検索キーが作成される(公報のインデキシング)。検索では、化合物の情報(構造式、名称など)を入力するが、入力情報は検索キーに変換されて、上記の公報インデックスに対して検索が行われる。

本検索方法は

- 検索キー抽出をテキストから行う点、
 - 図面内容やマーカッシュ形式の記載からは検索キーが作成されない点、
- 等から、通常のキーワード検索(以下、KW 検索と略記)に近いと言える。ただし、名称の異表記、商品名、CAS 登録番号なども合わせて検索キー化するので、検索もれは通常の KW 検索よりは少ないと

考えられる。

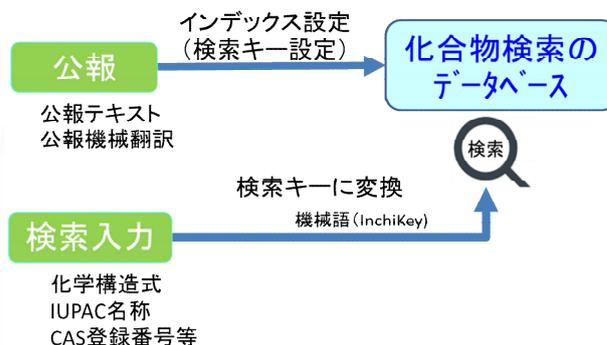


図2 PATENTSCOPE の検索キー設定と化合物検索

3.2 検索能力検証 KW 検索との比較-1 WO 公報、農薬の例

本報告の検索対象とした化合物 amidosulfuron の構造式を図3示す。本化合物を選んだ理由は、化合物が注目分野の農薬成分であること、やや複雑な構造を持っておりヒット件数がそれほど多くない、異表記が少ないため、解析が単純であると考えたためである。特許協力条約の出願(WIPO 出願、以下、WO 公報)の化合物検索を行った。化合物検索では、検索範囲は公報中の全テキスト部分(ALLTEXT)である。キーワード検索(以下、KW 検索)の検索領域もこれに合わせた。

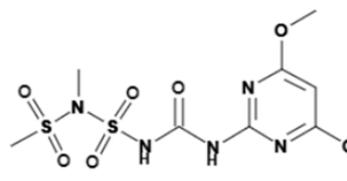


図3 amidosulfuron の化学構造式

図4に検索クエリー及びヒット集合の包含関係を示す。図中の集合 A,B,C は次の内容である。

A: 化合物検索でヒットするが、KW 検索

でヒットしないもの

B: 化合物検索、KW 検索ともにヒットするもの

C: KW 検索でヒットするが、化合物検索でヒットしないもの

(a) 検索クエリー

	Search Scope	Query	Hit patents
	Chemical compounds Search of "amidosulfuron"	CHEM:(CTTHWASMBLQOF R-UHFFFAOYSA-N) AND DP:([01.01.2000 TO 31.12.2017]) AND CTR:WO	2021
	Keyword Search of "amidosulfuron" ※	(ALLTXT:amidosulfuron*) AND DP:([01.01.2000 TO 31.12.2017]) AND CTR:WO	2338

※ amidosulfuronの前方一致検索

(b) ヒット集合の包含関係

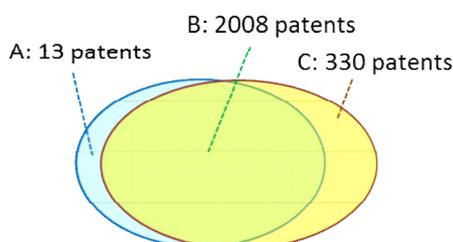


図4 amidosulfuronに関する検索クエリーとヒット集合の包含関係

図5に集合A、B、Cの出願受理国・出願言語を示す。図よりA、B、Cについて、つぎの点があった。

集合A

集合Aの13件の公報中では、amidosulfuron は、正規の文字列ではなく、不適切に区切られた文字列になっていた。不適切な文字列の内訳を図6に示す。これらは、KW 検索 amidosulfuron では、ヒットしない。化合物検索では、適切誤記修正が行われヒットしたことがわかる。

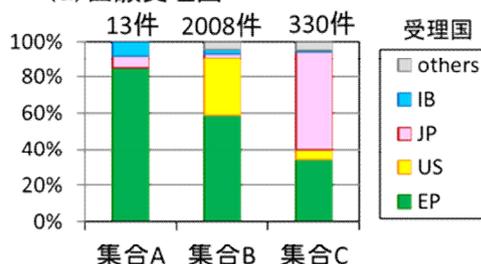
集合B

化合物検索のヒット集合の99%超がKW検索に含まれることを示す。図5(b)に示すように、出願言語はすべて英語・独語であり、検索対象の言語である。化合物検索でヒットしたことと矛盾がなかった。

集合C

化合物検索でヒットしなかったが、公報テキストを確認したところ、amidosulfuron の文字列は記載されていた。したがって、本来はヒットすべきである。図5(b)の公報言語では、化合物検索の指定言語である英語、独語の出願も存在する。出願言語の問題だけではないことが分かる。そこで、比較のため、出願言語が対象言語の英語であるUS出願について次節で検討を行った。

(a) 出願受理国



(b) 出願言語

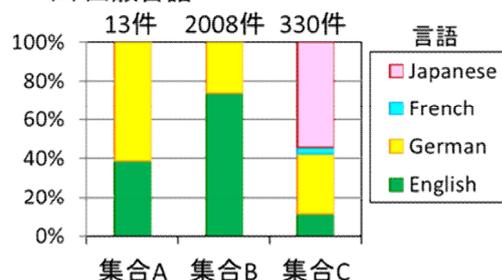


図5 集合A、B、Cの出願受理国・出願言語

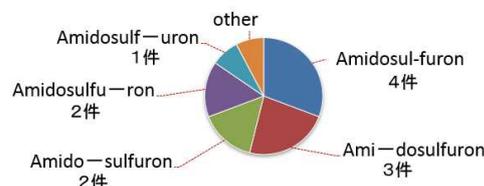


図6 集合A公報中の名称の誤記13件の内訳

3.3 検索能力検証 KW 検索との比較-2 US公報、農薬の例

前節3.2で出願国をUSに替えて同様な検討を行った。図7に検索クエリー、ヒ

ット集合の包含関係を示す。図中の集合 E,F,G は次の内容である。

- E: 化合物検索でヒット、KW 検索でヒットしないもの
- F: 化合物検索、KW 検索ともにヒットするもの
- G: KW 検索でヒット、化合物検索でヒットしない

(a) 検索クエリー

	Search Scope	Query	Hit patents
	Chemical compounds Search of "amidosulfuron"	CHEM:(CTTHWASMBLQOF R-UHFFFAOYSA-N) AND DP:([01.01.2000 TO 31.12.2017]) AND CTR:US	2101
	Keyword Search of "amidosulfuron" ※	(ALLTXT:amidosulfuron*) AND DP:([01.01.2000 TO 31.12.2017]) AND CTR:US	2441

※ amidosulfuronの前方一致検索

(b) ヒット集合の包含関係

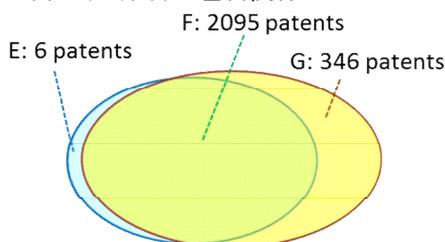


図7 amidosulfuronに関する検索クエリーとヒット集合の包含関係



図8 集合E, F, Gの出願言語

集合 E,F,G について、上記 3.2 節の集合 A,B,C と同様なことが分かった。

集合E

E の 6 件は、3.2 節の集合 A と同様に、公報中では、amidosulfuron が不適切に区切られた文字列にであった。不適切な文

字列の内訳を図9に示すが、図6の例と共通のものであった。

集合F

化合物検索のヒット集合の 99%超が KW 検索に含まれることを示す。言語も英語であり矛盾はない。

集合G

化合物検索でヒットしないものである。この中身は、amidosulfuron が記載されていたので、本来はヒットすべきものとする。公報の言語は英語であり、言語の問題ではないことが分かった。

3.4 化合物検索でヒットしない原因

ここで、集合 E、F、G について、

- 1) amidosulfuron の文字列が公報の明細書(Description)に存在するか、
- 2) 化合物検索キーの抽出が実際になされているか

を調べた。この調査は集合 F、G についてはそれぞれランダムに選んだ 20 件、集合 E は全件について行った。

結果を図10に示すが

- 検索キーの抽出はされていること、
- 明細書中に amidosulfuron の文字列は記載されていること、

を確認した。このことより、次の可能性があると考えられる。

- (i) 検索キーの抽出時のエラーによる、
- (ii) 検索キー抽出に使われたテキストデータが KW 検索のデータと異なる

集合 F のデータでは amidosulfuron の文字列を約 2000 件を間違いなく検索キー化できているので、(i)の可能性は低いと推定する。

出願時のデータがテキストデータでない場合には、化合物検索キーを抽出するタイミングがずれる可能性があり、上記(ii)となるのではないかと考える。

検索キー抽出の処理のタイミングを検討すれば、改善されるのではないかと推定する。対策が講じられ、修正されることを期待する。

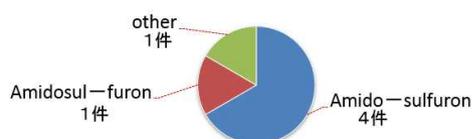


図9 集合Eの名称の誤記6件の内訳

	明細書中に文字列 amidosulfuron が存在する案件比率	化合物の検索キーが 抽出された案件割合
集合E	100%	100%
集合F	100%	100%
集合G	100%	100%

図10 明細書中の文字列存在比率、化合物検索キー抽出の比率

4. まとめ

PATENTSCOPEの化合物検索を農薬成分 amidosulfuron について検証した。KW 検索との比較では、次の点が分かった。

- i) 化合物検索でヒットした案件の99%超が KW 検索でもヒットした
- ii) キーワード検索でヒットした案件のうち、化合物検索でヒットしないものが約15%あった

結果を精査したところ、化合物検索でもヒットすべきことが分かったが、検索キーを抽出する公報テキストの情報が KW 検索と同じになるようにすれば解決すると思われる。この課題の解決により信頼性が向上することを期待する。

本報告では記載しなかったが、KW 検索のヒット件数のうち、化合物検索でもヒットしない案件は、Orbit でも相当比率存在し、PATENTSCOPE と同様な原因と考えられ、同様に解決を期待する。

5. 参考文献

- [1] 化学情報協会, ウェブサイト,
<https://www.jaici.or.jp/stn/caspat.html>
(Accessed on 2018-11-6)
- [2] 化学情報協会, ウェブサイト提供のテキスト,
https://www.jaici.or.jp/stn/pdf/text_chem1.pdf (Accessed on 2018-11-6)
- [3] 世界知的所有権機関(WIPO)ウェブサイト, PATENTSCOPE,
<https://patentscope2.wipo.int/search/ja/search.jsf> (Accessed on 2018-11-6)
- [4] 世界知的所有権機関(WIPO), “PATENTSCOPEの使い方”,
http://www.wipo.int/export/sites/www/about-wipo/ja/offices/japan/pdf/3_patentstscopex.pdf (Accessed on 2018-11-6)
- [5] 日本アグケム情報協議会オンライン研究会(2018/3/7-8)要旨集, “化学構造検索の比較検討”(2018)
- [6] 平成30年度PLASDOCオンライン研究会春季研修会(2018/6/15), 要旨集, “各種化合物検索ツールの性能評価”, (2018)
- [7] Magdalena Zelenkovska, “パテントスコープ(PATENTSCOPE)と特許情報検索のためのツール”, 特許・情報フェア&コンファレンス プレゼンテーション資料, pp.21-57 (2018)
- [8] 中央光学出版株式会社ウェブサイト, Orbit-Intelligence,
<https://www.cks.co.jp/home/product-1-2.htm> (Accessed on 2017-9-15)
- [9] 伊藤徹男, “PATENTSCOPEによるASEAN 特許”, Japio YEAR BOOK 2018, pp.168-177 (2018)