

化学情報データベースの作成者と利用者の両方の立場から

輔田 真 (42 期 1999 年卒)

1. 寄稿のきっかけ

2022 年 12 月に創薬有機化学研究室 周東智 教授の最終講義の案内が届きました。3 月中旬に札幌での開催ということで、早速スケジュールを調整して羽田～新千歳の航空券を予約しました。札幌を訪問するのは 20 年ぶりだったのでゆっくりしたかったのですが、年度末ということもあり日帰りでの参加となりました。

最終講義には幅広い年代にわたって多くの方が出席しており、元薬化学研究室のメンバーとも再会することができました。みなさん年を重ねてそれ相応に容姿は変わっていましたが、話しぶりや雰囲気などは学生時代とほとんど変わらず非常に懐かしく感じました。短い時間でしたが近況など色々情報交換をすることができ、非常に有意義な時間を過ごすことができました。

また、元薬化学研究室教授の松田 彰 先生にもお会いすることができました。松田先生には研究室に在席していた当時、そしてアメリカ留学の準備など大変お世話になりました。当時と変わらず現在も精力的に活動されているご様子でした。同窓会誌への寄稿を依頼されまして、少しでも恩返しになればと思い二つ返事で原稿の執筆を引き受けました。

東京に戻ってきて、翌週の月曜日に早速、会社のメールアドレス宛に芳香 ESSAY への寄稿依頼が来ました。依頼内容をよく確認すると 8,000～9,000 字にまとめてほしいと書いてあり…(冷や汗)。仕事が一段落した 4 月末から執筆を開始し、書いては消して、文字カウントを確認しては文章を追加して、なんとか完成にこぎつけました。気軽に目を通していただけますと幸いです。

2. 化学情報協会の紹介

本寄稿にあたり、松田先生からもご要望がございましたので、私が勤めています化学情

報協会について簡単に紹介をいたします。

私は博士課程修了後 2 年間フロリダ大学に留学し、帰国して以降は一般社団法人 化学情報協会 (JAICI) に勤めております。今年で 17 年目になります。化学情報協会の所在地は東京都文京区本駒込で、最寄り駅は山手線の駒込駅です。悲しいことに、友人や親戚に駒込に勤めていると言っても、それどこにあるの? と聞き返されることも多いです。30 個ある山手線の駅のなかでも知名度は高くはないほうだと思います。そんな駒込ですが、春の桜や秋の紅葉のライトアップで有名な六義園があります。園内のしだれ桜は全国的にも有名なもので、テレビのお天気中継や旅番組などでご覧になった方もおられるかもしれません。非常に手入れの行き届いた立派な庭園があり、自然豊かで素敵なおところです。入園料 300 円 (65 歳以上は 150 円) ですのでご興味のある方は一度足を運んでみてはいかがでしょうか。ちなみに、化学情報協会は道路 (本郷通り) を挟んで六義園の向かいに位置しています。

余談になりますが、ソメイヨシノも駒込が発祥の地と言われていますし、また、薬学部であれば 1 度は実験で使用したことがあると思われる駒込ピペットも駒込病院の院長により開発されたものだそうです。

本題に戻ります。化学情報協会の沿革ですが、1960 年代に米国化学会から Chemical Abstract (CA) 入力組織の設立の要請があり、1971 年に化学情報協議会が設立されたのがはじまりです。1977 年に社団法人化学情報協会に名称変更されました。その後 2011 年の法人移行を経て、現在の一般社団法人 化学情報協会となりました。

続いて化学情報協会の各部署を紹介いたします。まず「情報事業部」です。化学や医薬の分野の研究者で SciFinder (CAS SciFinder[®]) について聞いたことがある、または利用したことがあるという方も多いのではない

でしょうか。また、特許や文献を調べる際に STN (CAS STNext) を利用したことがあるという方もいらっしゃると思います。情報事業部ではエンドユーザー向けの科学技術情報の検索ツールである CAS SciFinder[®] や、情報専門家向けの科学技術分野の検索ツールである CAS STNext のサービスを提供しております。この他に CAS、FIZ Karlsruhe が提供するデータベースサービスの販売、普及、契約、ユーザーサポートを行っております。CAS SciFinder[®]、CAS STNext では随時情報のアップデートや機能の強化を行っており、近年では、CAS FORMULATIONS (医薬品・農薬・化粧品を中心とした製剤・配合情報を効率的に入手可能な製剤・配合情報に特化した検索サービス) や CAS Sequences (核酸・タンパク質の配列情報を検索するための CAS STNext 独自の機能) が追加されました。これらのデータベースの使用に関する講習会も適宜実施しておりますので、ご興味のある方はお問合せいただけますと幸いです。

また、「科学データ情報室」では結晶構造や物性データベース、分子設計・創薬支援ツールの提供しております。具体的には CCDC 製品 (有機化合物・有機金属化合物の結晶構造データベース)、ICSD (セラミックス、鉱物など、無機化合物の結晶構造データベース)、ASM Alloy Phase Diagram Database[™] (合金相図データベース)、NIST20 (質量スペクトルデータベース) などを取り扱っております。先日の最終講義で再会した某製薬会社に勤めている後輩から、結晶構造のデータベースを利用していると伺い、身近にユーザーがいることにうれしく思いました。

「情報技術部」では辞書サービス (JSD Pro/JSD)、機械翻訳サービス (JAICI AutoTrans、JAICI ProTranslator) を取り扱っております。独自の化合物表記翻訳を活用することで、特に中日翻訳などで誤訳率の低い翻訳を提供しています。特許調査をしていると、日本語、英語以外の言語、例えば中国語、韓国語、ドイツ語など、で記載された特許を読む機会が多くあります。Google 翻訳や DeepL 翻訳など無料の機械翻訳サービスもあります

が、これらは専門的な分野における翻訳、特に化合物名の翻訳の精度がいまひとつであることが多いように思います。より精度の高い機械翻訳をお探しの方は、ぜひ情報技術部にご相談ください。

「特許調査部」は、2005年に特許庁の先行技術調査の外部委託先として登録されました。それ以降登録調査機関として、有機化合物に関する先行技術調査 区分 30 (39 に分けられた技術分野の一つで、医薬・農薬・電子材料等の有機化合物特許) における、特許性の判断材料となる情報を調べて提供する業務を担当しています。

最後に、私が所属しています「知財情報センター (SHIPS)」の紹介です。SHIPS では医薬、化学、バイオの特許調査、文献調査のサービスを提供しています。「特許調査部」との違いは、特許庁ではなく、一般の企業、研究機関、大学、特許事務所などから依頼を受けて、調査を実施する点です。また、先行技術調査だけではなく、侵害予防調査、無効資料調査、技術動向調査など広範なタイプの調査を引き受けています。ありがたいことに、最近では大手の製薬会社様から継続的に調査をご依頼いただくことも多くなりました。はじめて調査を依頼されるお客様からは、調査の費用や納期についてお問い合わせいただくことがよくありますが、調査のタイプや技術分野、お客様が望まれる調査内容や成果物に応じて費用や納期は変わります。そのため、基本的にはまず打ち合わせ (コロナ禍以降はほとんど web 会議) を行い、お客様のご要望、ご予算、納期を伺ったうえで、それに対応したサービスを提供しております。

以上、化学情報協会では化学情報に関する多種多様なサービスを取り扱っておりますので、何かお困りごとや気になるサービスがありましたら、お気軽にお声がけいただけますと幸いです。

3. 化学情報協会での仕事

化学情報協会に入社してからしばらくの間は、Chemical Abstract (CA) のデータベース作成業務を担当していました。簡単に言うと、

公開された最新の日本特許の全文を読みこんで、その抄録や索引を作成する業務です。

最初はポリマー関連分野(樹脂組成物、成形体、フィルム、ゴム製品、繊維製品、塗料、紙類、洗剤など)の抄録・索引の作成を担当しました。大学院およびポストク時代には低分子化合物(大きな分子でもオリゴヌクレオチド程度)しか取り扱ってこなかったため、ポリマーにはまったく馴染みがありませんでした。低分子化合物とは違って、平均分子量やメルトフローレート(MFR)などの物性により多数の分子の集合体として取り扱う感覚になかなか馴染めませんでした。また、ポリマーの加工(成形、混練、押出、延伸など)に関する技術は参考書を読んでもイメージが掴みにくく、インターネットで公開されていた動画を閲覧して理解を深めました。不慣れなデータ入力作業、読みなれない特許、新たな技術分野の習得と三重の苦労がありましたが、現在同じ部署で一緒に働いている M 先輩が当時指導者として熱心に教育してくださり、無事乗り越えることができました。M 先輩に直接ご指導いただいたのは非常に幸運だったと思うし、今でも大変感謝しています。

入社から数年後に担当する技術分野を拡大することになり、半導体、光学、磁気関連分野の抄録や索引の作成も行うことになりました。技術分野が薬学とはどんどん離れていき、学生時代に学んだことを仕事に生かせないもどかしさはありませんでしたが、新たな分野を開拓するのは新鮮で刺激的なものでした。

またこの時期に抄録・索引の作成業務と並行して、CA とは異なるデータベースのデータ入力を担当する機会に恵まれました。この業務を遂行するために早朝 6 時から Webex で CAS とのミーティングに参加したり、2 週間ほどアメリカで研修を受けたりと大変ではありましたが、今となっては良い思い出です。これまでのような新しい技術分野について学ぶ必要はありませんでしたが、有機化学のパズルを解いているような特別な業務でした。

入社から 10 年ほどして転機を迎えまして、知財情報センター(SHIPS)に異動になりました。これまでは特許のデータベースを作成する

立場でしたが、そのデータベースを利用して特許や文献を調査する立場になりました。データを入力する作業と、それを検索する作業に共通する部分はあるものの、効率的な調査を行うためには専門の検索技術を身につける必要があります、一から検索の勉強をやりなおしました。

また、特許調査を行う上で特許法を含む知的財産法についても学ぶ必要があります、3 か月間、独立行政法人 工業所有権情報・研修館(INPIT)の研修に参加しました。薬剤師の国家試験や大学院の入試の時以来、必死で勉強したように思います。不幸にもこの時期に尿路結石になりまして、土曜日の深夜に救急車で駒込病院に運ばれました。子供の頃に歯医者に通ったくらいで、冬に大型の雪像を作っているときに事故でスコップが頭に刺さっても病院に行かなかったくらい体の丈夫さには自信があったのですが、よもや自分で救急車を呼ぶことになるとは思いもしませんでした。石がなかなか排出されず、あまり痛み止めも効かなかったので、唐突に襲ってくる激痛に恐怖しながらの生活が 3 週間くらい続きました。INPIT の研修のレポート作成や試験の準備と重なって非常に辛かったです。結石の原因は色々あるようですが、このとき以来、ほうれん草やチョコレートのようなシュウ酸を多く含む食品を控えるようにしています。なんとか INPIT の研修を修了することができ、実務を通して経験を積む毎日が続いております。

4. データベースの作成者としての観点

化学情報協会に入社してから 10 年間ほど担当していた Chemical Abstract(CA)のデータ入力(特許の抄録、索引の作成)について簡単に紹介いたします。抄録、索引といわれても、どのようなものかイメージしにくいかもしれませんが、抄録は特許の発明の特徴や長所(従来技術より優れている点)を端的に文章でまとめたものです。一方で索引は、特許の発明に関わる重要な要素(化合物、用途、プロセス、物性など)を参照できるようにまとめたものです。仮想の例として下記の特許「ポリエステルフィルム」の抄録と索引を作成すると以下のように

なります。

特許「ポリエステルフィルム」

- ・ 請求項に、生分解性ポリエステルとリン系難燃剤を含むフィルムが記載
- ・ 実施例に、ポリ(L-乳酸)と TPP(リン酸トリフェニル)からなる押出フィルムの難燃性 UL94 について記載
- ・ 明細書中に、課題として生分解性フィルムの難燃性が乏しいこと、用途として包装に適していることが記載

<抄録>

The title film, suitable for packaging, comprises biodegradable polyesters and a phosphorus fireproofing agent. Poly(lactic acid) was melt-kneaded with triphenyl phosphate (TPP), extruded through a die, and stretched to give a film with fire resistance V-0 (UL94).

<索引>

Poly(lactic acid)、Polyesters、TPP、Fireproofing agents、Fire-resistant materials、Biodegradable materials、Packaging films
※各索引にはさらに説明やロール(化学物質の役割)が付与される

実際の CA のデータ入力では、さらに技術分野に応じたセクションの付与や補遺語(原報の表題や抄録中の重要な用語)の追加などをおこないます。抄録は入力できる長さに制限があるため、多数の部品からなる装置、多数の工程からなる製造方法、従来技術との違いが不明瞭な発明の特許などは、抄録の作成が大変です。また、上記の仮想の例では索引が 10 個程度ですが、請求項や実施例の多い特許では数百を超える索引が必要になる場合もあります。

5. データベースの利用者としての観点

データベースを利用する立場になってわかったことは、「必要な情報を過不足なく検索により見つけ出すのは難しい」ということです。先ほどの仮想例で登場した難燃剤の「リン酸トリ

フェニル」を例に考えてみたいと思います。もし、リン酸トリフェニルが記載されている特許のみを漏れなく検索したい場合、どうすればよいでしょうか？

検索方法

パターン 1: 「リン酸トリフェニル」を検索

パターン 2: 「リン酸トリフェニル」の同義語(例えば、トリフェニルホスフェート)を検索

パターン 3: 「リン酸トリフェニル」の略語 TPP を検索

パターン 4: 「リン酸トリフェニル」の組成式(例えば、OP(OC6H5)3)を検索

パターン 1 の検索は真っ先に思いつく検索で、一見問題が無いように思われます。しかしながら、この検索語をそのまま検索すると「亜リン酸トリフェニル」がヒットする可能性があります。またパターン 3 の検索を実施した場合、同じ略語をもつ「テトラフェニルポルフィリン」や「チアミンピロリン酸」がヒットする可能性があります。このような単語の検索にはどうしても限界があり、最終的には検索でヒットした特許をすべて査読(スクリーニング)して、当たりかどうかを判断する必要があります。また、パターン 1~4 の検索をすべて実施したとしても、リン酸トリフェニルの構造式が図示されているものや Markush 構造式のような一般式で記載されているものはヒットしません。

通常の特許や文献の調査では、このように 1 つの単語の検索だけで終了するということはほとんどありません。複数の概念の検索語、特許分類、統制語(CA の統制語や MEDLINE の MeSH タームなど)、CAS RN®などを色々組み合わせ、調査の目的に合致したノイズの少ない特許や文献の集合の作成を目指します。先ほどのリン酸トリフェニルの例について、もし 1 件も漏れない調査が必要という状況であれば、パターン 1~4 に加えて CAS RN®などを利用した追加の検索が必要になるかもしれません。一方で、目的の特許が数件でも見つければよいという状況であれば、パターン 1 の検索だけで事足りる可能性があります。

特許や学術文献を検索するためのさまざま

な検索システムやサービスが提供されています。必要な情報を過不足なく検索するには、これらの検索システムや収録されているデータベースの内容についてよく知ることも重要です。Google や Yahoo など無料の検索サイトは手軽に利用できて非常に便利なので私も日常的に使用しています。これらの検索サイトで得られる情報は内容が不正確であることや、情報源がはっきりしない場合があるため、実際の特許・文献の調査では専用の検索システムを利用することが多いです。複数の種類の専用の特許検索システムを使用して見たことがありますが、基本的な検索機能はどれも類似しており、近傍検索のような特殊な検索機能、主要国以外の特許の収録状況、検索結果の解析機能などにおいて、各種サービスで違いがあるように思います。

ちなみに、私が抄録や索引を作成していた CA は、一次情報(特許や学術文献の情報)を加工した二次情報(抄録、索引など)のデータベースです。先ほどのリン酸トリフェニルのように、トリフェニルホスフェート、TPP、OP(OC6H5)3、図示された構造式のように記載に揺らぎがあっても、同じ化合物であれば同じ CAS RN[®]が付与されます。また、類似の用途やプロセスであればそれぞれ共通の統制語が索引されます。そのため CAS RN[®]や統制語を利用することで、表記のゆらぎの影響を受けずに目的の特許や文献を検索することができます。ただし、これらの情報は人手で入力されるため、データ入力の担当者によってバラツキがあります。また、品質チェックなどの対策を講じていますが、入力ミスを完全に避けることはできません。一方で、一次情報のデータベースであればデータ入力のばらつきや入力ミスのような問題は生じませんが、図示された構造式や PDF 化された表中の文言を検索することは容易ではありません。また、特許によってはテキスト部分が文字化けしている場合や、わざとかどうかは不明ですが 1(数字のいち)が 1(小文字のエール)と入力されている場合があり、正しい検索語を使用しても検索でヒットしないことがあります。

以上のことから、調査の内容に応じて適切

な検索システムを選択すること、そして欠点を補うことができるように複数の検索システムを使用することが有効であると考えています。最終的には、調査の費用や納期を考慮したうえで、検索の方法を決定することになります。

6. コロナ禍での生活

2019 年の 2 月に新型コロナウイルスの流行が始まり、生活は一変しました。特に初年度の緊急事態宣言が発出されている間は、原則在宅勤務となり、毎日人で溢れかえっている東京でさえも一時期ゴーストタウンのようになっていました。私は自宅が会社の徒歩圏内ということもあり、許可をいただいて入社して通常どおり勤務を続けていました。在宅勤務者のパソコンのリモート接続トラブルの対応や、他部署の観葉植物の世話もしました。お客様と対面で打ち合わせをすることがなくなり、ほとんどすべて web 会議に切り替わりました。最初は機器の接続や、お客様との会話や相槌のタイミングが合わないなどで苦労しましたが、今ではかなりスムーズに一連の作業を行うことができるようになりました。web 会議は非常に便利で仕事の効率もかなり上がったように思います。ただ、意思疎通のしやすさを考えると、時間さえ許せば対面のほうが自分には合っているかなと感じました。つい先日、新型コロナの感染症法上の位置づけが 5 類に移行しました。東京では 3 月あたりから海外からの旅行客をかなり目にするようになり、またスポーツや祭りなどイベントも通常どおり開催されるようになって、コロナ前の生活に近い状態に戻つつあるように思います。

7. 近況

前記しましたように 40 歳になってはじめて救急車を経験し、2 年ほど前から老眼が急速に進んで、暗いところや近いところで字を読むのが大変になり、老化や衰えを実感する日々です。同年代の仲間うちでは健康や介護に関する会話が増えました。さらには、年が明けたと思ったら、もうゴールデンウィークを過ぎており、気づいたら年末になっているのではと思うくらい時間の進み方が早く感じます。

化学情報協会に入社してからは薬学とは異なる技術分野を担当してきましたが、SHIPSで特許や文献の調査をはじめようになって大学院時代に学んだことが仕事でとても役立っています。松田先生や周東先生のもとで核酸化学や創薬化学を学ぶことができたのは本当に幸運でした。ただ、当時は合成系に注力して生物系の勉強をさぼっていたので、もっとバイオ関連の分野も貪欲に学んでおけばよかったと後悔しているところです。

最後に、市川先生(北海道大学大学院薬学研究院教授)や阿部先生(名古屋大学大学院理学研究科教授)をはじめ、当時、学部生、大学院生として一緒に薬化学研究室に在席していた方が、現在、教授や准教授とし

てアカデミックの世界で頑張っておられることに非常に誇りに思います。特許や文献の調査の仕事をしていて、特許の発明者や論文の著者にみなさんの名前を見つけると自分も頑張らねばと大変励みになります。医薬、バイオ系は昔から競争の激しい分野ですが、CRISPR遺伝子改変技術が開発されて以降、より一層技術の進歩が早まっている印象です。厳しい研究環境とは思いますが、さらなるみなさまのご活躍を期待しております。

同窓会 HP:2023年6月27日公開
2023年8月23日追記