

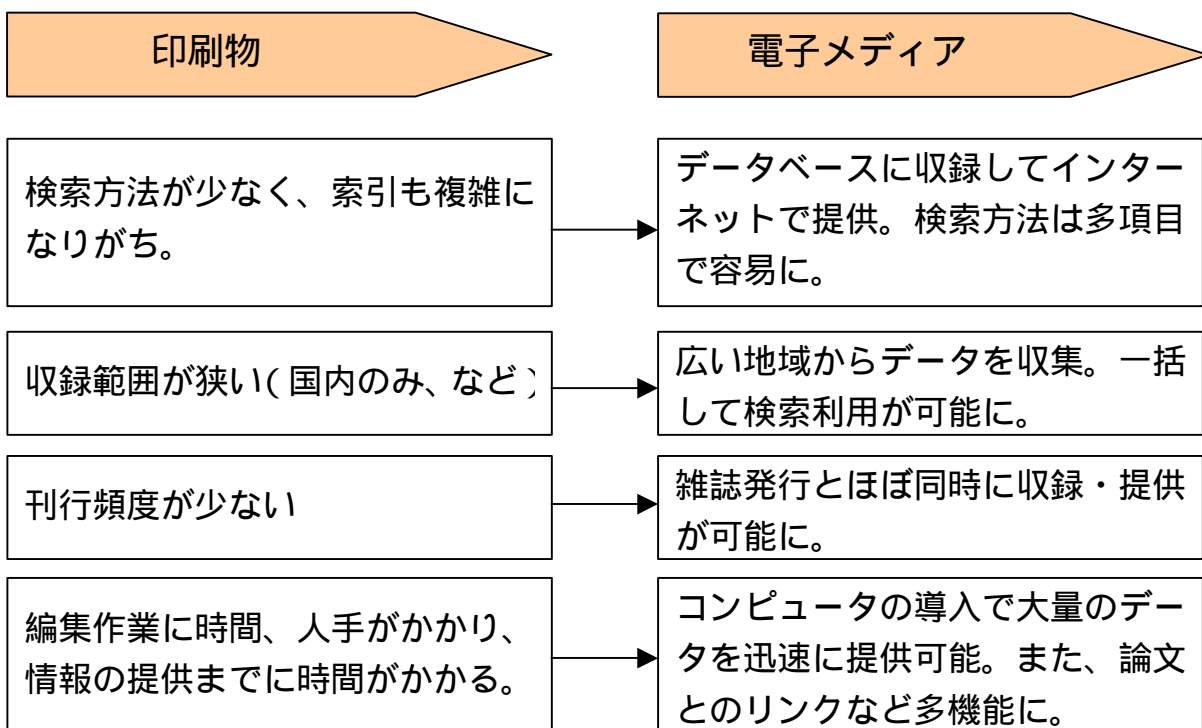
3.1 データベースを活用しよう

東北大学では、種々のデータベースを導入していますが、これらを利用するとどんなことができるのでしょうか。個々のデータベースの詳細は後の章で詳しく説明しますが、まずデータベースを利用する上での特徴を理解しておきましょう。

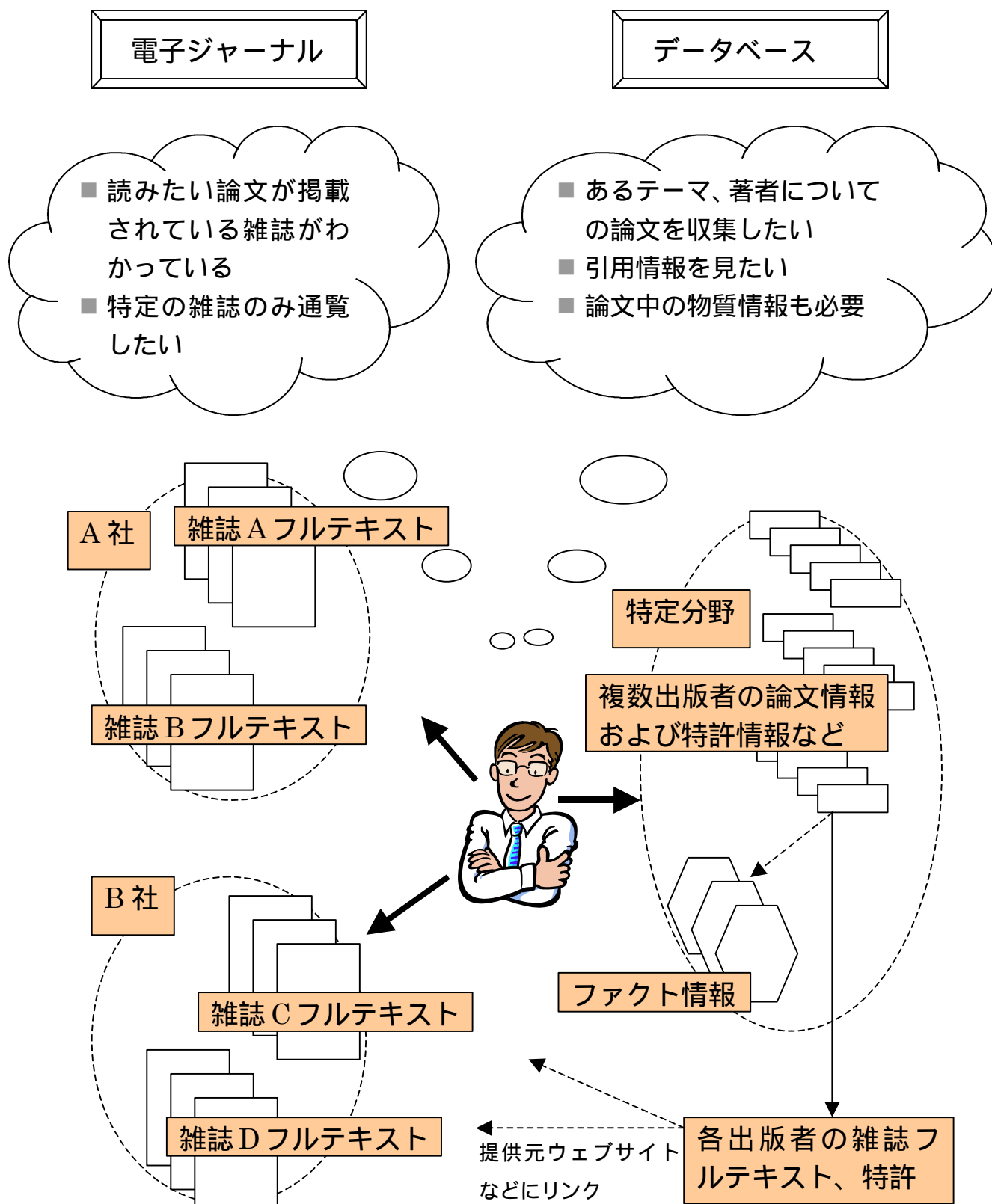
3.1.1 データベースの変遷と電子ジャーナルとの使い分け

現在、大規模なデータベースとして成長しているものの多くは、従来から冊子体で長く刊行されてきた歴史があるものです。蓄積された大量のデータを、現在の情報処理技術と発達したネットワークによって、簡便にかつ早く提供することができるようになりました。

前章で説明した電子ジャーナルも、著者や、論文タイトルなどから検索できる機能がついており、かなりデータベース的な要素を持っています。さらに最新の情報を常に配信されるようにしておけるアラートサービスもあり、論文作成のための情報収集にはこれで十分と思われるかもしれません。実際、前述のようなサービスを利用するだけでも、すべてに目を通す時間が確保できないほどの論文情報が手に入ります。電子ジャーナルとデータベースの使い方の特徴についても理解しておきましょう。



図表 3.1-1 印刷物から電子メディアへ



図表 3.1-2 電子ジャーナルとデータベースの使い分け

3.1.2 データベースの種類

身近にはさまざまな「データベース」がありますが、ここでは、東北大学で利用できるデータベースを中心に紹介します。種類としては、大きくわけて、どんな論文や特許が発表されたのかを知るための文献データベースと、過去の実験などから導かれた物質の計測値などを集めたファクトデータベースがあります。

(1) 文献データベース

雑誌論文、特許、学位論文など、これまでに発表された論文の書誌情報(タイトル、著者など)を蓄積しているものです。自然科学系は早くからこのような資料の需要があったため作成の歴史も長く、特に化学の分野は19世紀の中頃から冊子体のかたちで本格的な作成が始まっています。

例えば、『Chemical Abstracts』は、刊行開始は1907年ですが、そのルーツは、19世紀の半ばまで遡ることができます。長く冊子体でのみ提供されてきましたが、コンピュータやネットワークの発達、普及に伴い、提供メディアをCD-ROM、オンラインなどに広げながら提供されています。本書でも紹介している『SciFinder Scholar』はその全データを検索利用でき、操作も簡便な最新の提供形態です。

また、医学系のデータベースである『MEDLINE』も、『Index Medicus』(1879~)という冊子体がもとになっています。『Chemical Abstracts』もそうですが、こちらも提供形態はさまざまに広がり、データベースでの利用が主流となった現在でも冊子体の提供も継続されています。

さらに、『Biological Abstracts』は、同名の冊子体は1926年刊行ですが、その前身である『Botanical Abstracts』までさかのぼると1918年から刊行されています。

これらの例のように、多様な利用方法に対応できるよう、データの蓄積とともに提供形態も変わってきています。注意しなくてはならないことは、もともとのデータが同じでも、提供される形態が異なれば、検索方法および検索結果が異なることがある、ということです。冊子体の場合、CD-ROMで利用する場合、オンラインサービスで利用する場合、また、同じオンラインでも検索ソフトが異なれば事情が異なります。使い慣れたデータベースでも新しい形態でサービスが提供された場合は、その特徴をよく理解しておきましょう。

(2) ファクトデータベース

自然科学系で重要な情報である各種の数値データ、物性データなどについては、従来から事典、ハンドブック類などで情報が提供されてきました。個々のデータは、研究成果として雑誌論文などで発表された数値や、計測専門の機関などによって公表されたものです。中には後からの検証により数値が訂正される場合もあります。従来から冊子体で刊行されてきたものは、収集や更新が難しいこともあって、特定の狭い範囲を扱っていたり、主な数値のみを挙げるダイジェスト版であったりします。基本的な情報はそれらに掲載されている場合も多いので、それぞれに版を重ね、重要な参考図書として利用されてきました。

現在では「網羅的であること」「最新であること」「Abstract あるいはフルテキストから関連する物質情報が即参照できること」が当然のように求められており、これらの要求に応えようとするデータベースが登場してきました。例として本書では『CrossFire』を紹介しています(3.7 参照)。

ほかに各研究機関においても、その成果である数値データをデータベースとして公開している場合もあり、特定の分野のみが収録対象ではありますが、貴重な情報源として利用されています。



(3) 東北大学で利用できるデータベース

自然科学系の主なデータベースは以下のとおりです。調べたい内容によって、データベースを選択します。中には利用に先立って申請および費用負担が必要なものがあります。自分の所属する研究室から利用できるかどうか確認しておきましょう。

分野	データベース名および分野・収録年代など	収録資料・特徴など
全般	Web of Science ■ 自然科学分野 (1945-)ほか	学術雑誌論文(日本語雑誌も、英文抄録があり、収録対象として選定されていれば含まれます) ■ あるテーマ(著者)についての論文検索 ■ 個々の論文の引用に関する情報 3.2 参照
	Current Contents (Ovid Online) ■ 全分野(1993-)	学術雑誌論文・新刊書・会議録 ■ 各雑誌の目次情報を一覧可能 ■ 抄録が読めるものもある 3.8 参照
	inside web ■ 全分野(1993-) (英国図書館所蔵資料)	学術雑誌論文・会議録 ■ 所蔵されていれば、図書館を通じて、あるいは個別に契約している研究室で文献取り寄せが可能 3.8 参照
	雑誌記事索引 (NDL-OPAC) ■ 全分野(1948-) (国立国会図書館所蔵資料)	国内発行雑誌の記事 ■ 所蔵されていれば、図書館を通じて文献取り寄せが可能 ■ 自然科学系記事の収録は1975年以降
生命科学	MEDLINE (Ovid Online) ■ 医学・生物学(1951-)	学術雑誌論文(日本語雑誌も収録対象として選定されていれば含まれます)、図書 他に作成元である NLM から提供されている『PubMed』が無料で利用できます。 3.4 参照
	医中誌 Web ■ 医学・歯学・薬学 (1983-)	国内で発行される医学系の学術雑誌論文 ■ シソーラスが利用可能 3.5 参照

	Biological Abstracts およびRRM (Ovid Online) ■ 生物学(1985-)	学術雑誌論文、テクニカル・レポート、レビュー、会議録 ■ シソーラスが利用可能 3.6 参照
化学・物理	SciFinder Scholar ■ 1840- ■ 化学・物理・医学・薬学	学術雑誌論文、会議録、特許、学位論文、テクニカル・レポート ■ 物質情報は、「化学物質辞書ファイル(CAS 作成)」から収録。 ■ 反応情報は「反応情報データベース(CAS 作成)」から収録。 ■ 『MEDLINE』も同時に検索可能。 3.3 参照
	CrossFire Beilstein/Gmelin ■ Beilstein(有機) 1771- ■ Gmelin (無機) 1772-	化合物情報および付随する文献情報 ■ 化学構造式を図示しての検索、語句からの検索、物質の物性を限定しての検索などが可能 3.7 参照
工学・物理	Inspec ■ 物理・電気・電子(1969-)	学術雑誌論文、図書、テクニカル・レポート、会議録 ■ 全データに抄録付き 3.8 参照
	Ei Compendex (CD-ROM) ■ 工学(1970-) (工学分館所蔵 1993-)	学術雑誌論文、図書、テクニカル・レポート、会議録 ■ 応用工学領域を広くカバー 3.8 参照
数学	MathSciNet ■ 数学(1940-)	学術雑誌論文、図書、会議録 ■ American Mathematical Society 発行の『Mathematical reviews』のオンライン版 3.8 参照
	Zentralblatt MATH ■ 数学(1868-)	学術雑誌論文、図書、会議録 ■ European Mathematical Society 発行の同名の抄録誌のオンライン版 3.8 参照

図表 3.1-3 データベースの使い分け

コラム 使えるデータベースは毎年変わる？

学内では、全学で費用負担し、個別の利用者に負担を求めないデータベースと、それを必要とする研究室で費用を分担して負担し、利用しているデータベースなどがあります。

前者の代表的なものは『Web of Science』で、東北大学の構成員であれば、教官のみならず学生でも職員でも個別の負担なしで利用ができます。

後者の代表は現在のところ『SciFinder Scholar』です。年ごとに利用希望を募り、図書館で一括契約し、費用はその年利用する研究室で平等に負担します。

こうした学内でのデータベース選定に関しては、学内の各部局からの代表者によって構成される委員会で決定されます。

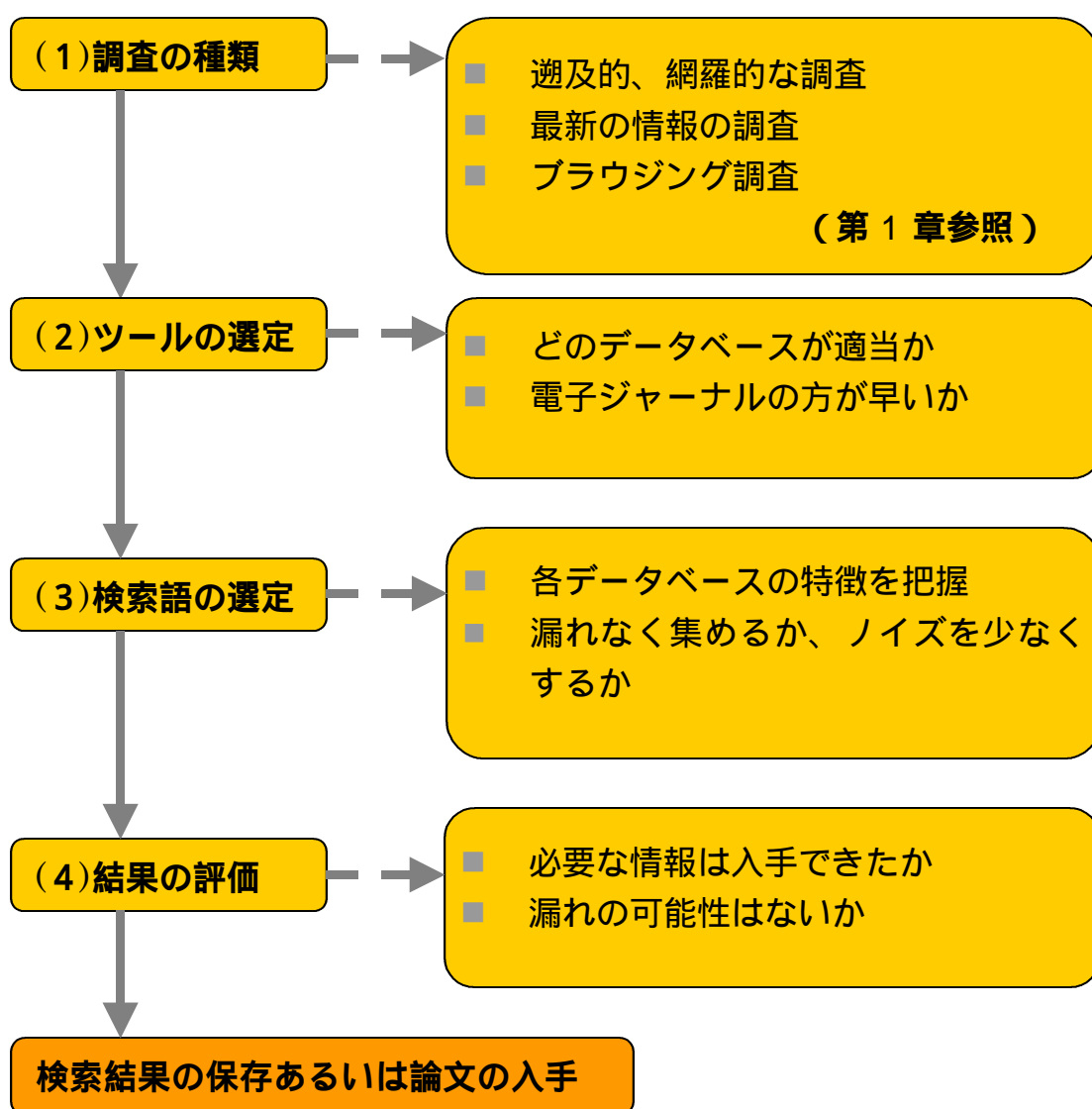
したがって、その年の学内の意見により、入れ替わりはありえます。新しく、かつ多くの教員が必要だと認めた場合は全学での費用負担として、導入される可能性がありますし、逆に不要と考えられるものについては利用停止となることがあります。

その年の利用については、附属図書館から広報されますので、利用、あるいは申請の呼びかけがあった際には、昨年との変更点などにご注意ください。申請の仕方、マニュアル、新しいデータベースのトライアルなどの情報も附属図書館のウェブサイトに掲載されます。操作方法についての問い合わせ先も掲載されていますので、ぜひご覧ください。

<http://www.library.tohoku.ac.jp/dbsi/>
(学内で利用できるデータベース一覧)

3.1.3 利用上手になるために

これらのデータベースを実際に使いこなすためには具体的にどんなことをしたらよいのでしょうか。効率の良い利用のため、事前に検索の手順を考えておくことを、検索戦略をたてる、といいます。以下に、まず手順の概略をまとめ、それぞれの段階での具体的なポイントを順を追って説明します。



図表 3.1-4 検索の流れと検索戦略

(1) 調査の種類

これから自分が行う調査は、どの範囲の資料をどれくらい集めればよいのか、集めた結果どうするのか、などを考えた上で始めた方が無駄がありません。第1章で紹介したように、網羅的な調査が必要なもの、最新の情報を集めたいものなど、研究段階に応じた調査方法があります。

また、「何の情報を求めているのか」を明確にしておく必要があります。文献情報のみでよいのか、物質に関する情報なのか、特許は含むのか、など必要とする情報によって探索方法も変わってきます。

(2) ツールの選定

データベースを利用するにあたって、注意が必要なことは以下のとおりです。

■ 収録範囲（分野）は合っているか

利用しようと思うデータベースは、自分が探している分野を収録対象としているかどうか、です。自然科学系の各領域について境界領域と呼ぶべき分野が増えてきたことに対応し、大規模データベースでは収録範囲をかなり広くしていますが、得意分野はそれぞれにあります。

■ 収録範囲（年代）は合っているか

データベースによって、いつからのデータを収録しているかが異なります。場合によっては、データベース収録対象になっていないこともありますので、冊子体で利用できるツールがないかどうかカウンターにご相談ください。

■ 収録資料の種類は合っているか

会議録、特許、学位論文などは、データベースによっては収録対象としていません。それぞれ個別のデータベースも存在しますので、後の各章を参照してください。

(3) 検索語の選定

データベースが提供され始めた当時は、コマンドによる検索が主流でした。データベースの構成に加え、各コマンドの使い方、検索演算子の指定の仕方など、あらかじめ学習しておかなくてはならないことがたくさんありました。料金も高額かつ従量制だったため、データベースを利用する場合は、検索を実行する前にどのようなコマンドをどのような順番で入力するのが適当かを考えなくてはならず、ここから「検索戦略をたてる」という表現がうまれました。

現在は、ブラウザでの提供が主となり、またそうでない場合でも、特に事前の学習なしで利用者自身が検索することが可能になっています。しかし、インターフェイスが親しみやすくなっただけで、データベース内部の複雑さは増しているということもあります。同義語、類語などの自動処理の有無、上位・下位語などの概念の有無などは、検索語の選定時に重要ですが、データベースによりさまざまです。ここまでは直感だけではわかりません。面倒でもマニュアルやヘルプ画面を参照する必要があります。

各データベースから、必要とする情報を的確に取り出すには現在でも技術が必要であり、その能力を身に付けるための「情報リテラシー」教育が広がっています。戦略、という語は現在は大げさな印象がありますが、検索に先立って図表 3.1-4 のような事項をおさえ、ある程度プランを練ることは効率的な検索を行う上で必要なことです。

コラム 近接演算子とは？

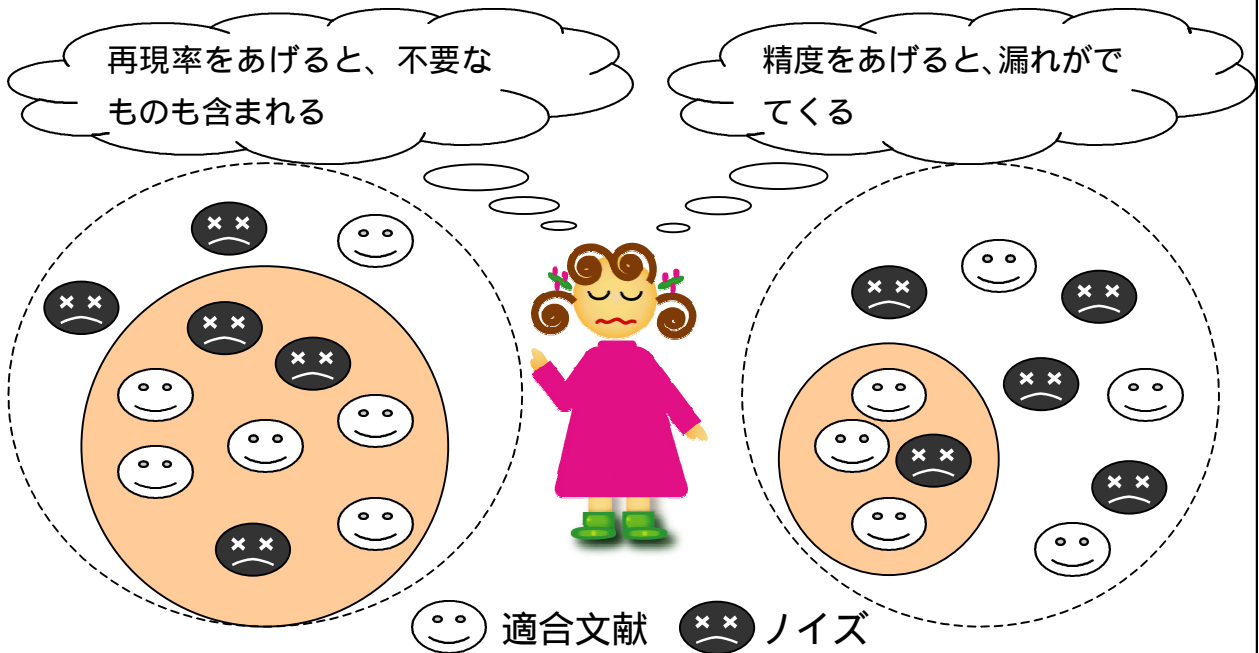
近接演算子は、複数の検索語間の関係を指定するために使います。すべてのデータベースで利用できるわけではなく、利用可能な指定の種類や記号の使用方法も異なりますが、主として以下の機能を持ちます。本書では利用の一例として、3.7の「AutoSearch」のコラムで紹介しています。

- 語同士が隣接している
- 語同士が指定した範囲の中に出現する（5語以内に出現する、同一センテンスの中にある、など）
- 語同士の出現順序を指定する

コラム データベースにおける再現率と精度

データベースから、自分が必要とするデータを検索する場合、もともとデータベースに登録されていた適合文献をどれだけ検索できたか、を表すのが再現率です。「漏れなく網羅的に」検索したい場合は、再現率が高くなるような検索をする必要があります。具体的には、キーワードとしていくつもの同義語、類語などを入力する、必要に応じて関連語からも検索するなどが考えられます。この場合の問題は、いわゆる「ノイズ」が多くなることです。情報としては関連しているけれども自分が求めていた情報とは異なるものが多くでてきてしまいます。

結果として表示された集合の中に、自分が求めていた情報がどれだけ含まれていたかは「精度」であらわします。一般的に「再現率」をあげようとすれば「精度」が下がり、「精度」を上げようとすると「再現率」が下がり、必要な情報が漏れてしまう可能性があります。



自分が行う検索が、どちらを重視したほうがよいのかは、その時々の研究の段階によるでしょう。漫然とキーワードを入力するのではなく、検索で集めたい情報の範囲を意識しながら検索できるようになると、あとからまた同じような検索をしなくて済みます。

(4) 結果の評価

最近のデータベースは、マニュアルを読まないとも出せない、というほどのものではなく、どれもわかりやすい画面構成となっています。だからと言って、「とりあえず思いついたキーワードを入力して、何かしら結果が返ってきたから使いこなせている」と思ったのでは十分ではありません。入力したキーワードに対して、なぜその結果が返ってきたのか説明できなければ「その検索で十分な結果が得られた」とは言えないでしょう。

数多くのデータベースを使っている人ほど、「各データベースの違い」について敏感になるようです。同じような結果が返ってくるべき場面で異なる結果となったとき、初めて「入力すべきキーワードに何か規則があるのか」「収録されているデータに違いがあるのか」「使用している記号に問題があるのか」などについて疑問に思うことになります。

少し使ってみて様子が出てからでよいので、マニュアルには目を通すようにしてください。自分が行っていた検索方法で正しかったのか、あるいは「漏れ」がある可能性があるのか、どうすれば最も効率的に利用できるのか、が明確に記載されています。さらになにか問題があった場合の問い合わせ先（ヘルプデスク）も記載されています。

オンラインのヘルプ画面も便利ですが、検索の途中だと画面が重なったりして見づらい場合があります。新しいデータベースを使い始めるときは、いつでも手にとれるようにマニュアルを手元においておくことをお勧めします。

コラム シソーラスとは？

データベースで検索を行う場合、どのようなキーワードを入力したら最適な結果が得られるか、は非常に大きな問題です。的確な検索を可能にするため、関連する語、同義語、上位概念語、下位概念語などを関連付けし作成された表をシソーラス（thesaurus）と言います。例えば医学関係のデータベース『MEDLINE』では、キーワードとして「cancer」と入力すると、「Neoplasma」という最上位概念に誘導し、そこから例えば「患部の指定による」「骨」「頭蓋」などと下位の概念に進むことができます。これを利用することで必要に応じ、検索の幅を広げたり限定したりすることが可能です。このように関連づけされた語を「統制語」といいます。

(5) 思うような結果がでないとき

さて、各データベースの特徴も把握したし、マニュアルも読んだし、入力したキーワードも問題がないはずなのに思うように結果がでない場合があります。このようなときは、以下のことを試してみましょう。

■ 別の角度から検索してみましょう

選択したキーワードが、予想した形では収録されていないこともあります。マニュアルどおり入力していても、例外も多々あります。判明している事項の中で他に利用できそうなキーワードがあれば、そこから再検索し、絞り込んでいくと見つかる場合があります。

また、例えば物質に関する情報を入手したくて『CrossFire』を利用したとします。このときうまく結果がでない場合、『SciFinder Scholar』で検索し、その結果をもとに再度『CrossFire』を検索するという使い方もあります。もちろんその逆もありますし、別のデータベースとの組み合わせもあり得るでしょう。一つのデータベースのみで検索するのではなく、他のデータベースと補い合うことで結果がでることもあります。

手がかりが少ない時は、一般的なサーチエンジンによりヒントを得られないか試してみましょう。探している著者、あるいは物質に関する情報について、関連する研究機関などが情報をまとめているページにより、有用なキーワードが見つかったり、入手していた情報の誤りなどに気づいたりします。それを利用して、再度データベースで検索してみます。

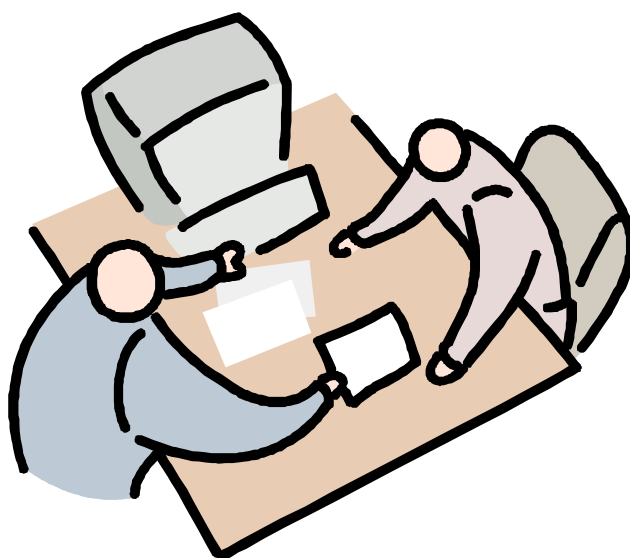
■ 他の有料のデータベースも試みましょう

東北大学では、多くのデータベースを利用者の直接負担なしで利用できるようにしていますが、それ以外でも有用なデータベースはたくさんあります。『SciFinder Scholar』や『CrossFire』は現在は利用する研究室ごとに負担があり、附属図書館を窓口として全学からの申し込みを受け付けています。また、これらは附属図書館、分館内の専用パソコンからは負担なしで利用できますので、学生でも自由に使うことができます。自分がもとめる情報は、これら以外のデータベースに収録されている可能性もあります。

有料の代表的なサービスとしては、『Dialog』や『STN』、『NACSIS-IR』、『JOIS』などがあり、それぞれ多数のデータベースを提供しています。しかし、現在のところこれらのサービスを利用するには事前に利用申請が必要で、利用の都度料金がかかります。図書館（図書室を含む）によっては、それらを校費で利用できるようIDを取得しているところもありますので、問い合わせてみてください。もし、利用してみて自分の研究、あるいは研究室全体にとって非常に有用な場合は、研究室単位、あるいは個人単位で利用契約を結ぶとよいでしょう。

■ 図書館員に相談してみましょう

ある程度やってみて次にどうすべきか、と迷ったときは、一度図書館員に相談してみてください。「このような情報が欲しくて、これらの検索をしたが結果が思わしくない」ことを伝えれば、他にどのような探し方があるか担当者の方で調査します。そうすることで、自分の時間を別のことに振り向けることができますし、担当者から連絡があれば、またそこから検索を始めることができます。



コラム データベース利用時の注意あれこれ

■ 同時アクセス数

契約内容により、学内から同時にアクセスできる利用者数が決まっています。例えば、「5」と設定されている場合は、6人目の人が利用を開始しようとする以下のようなメッセージがでて利用できません。

例) 『SciFinder Scholar』の場合

“Login Failed: Too many concurrent sessions for this login id.”

利用中の誰かが終了すれば、すぐ新規に利用できる所以少し時間をおいて試してください。同時アクセス数は、実際の利用状況を見ながら、利用に支障をきたすことのないよう設定されていますが、時期、あるいは時間帯によっては若干混み合うこともあります。ほかの利用者のために、利用が終わったらすみやかに終了操作（データベースにより、「Exit」「Logout」「Logoff」などと表現されます）をしてください。

■ セッション

一度検索をはじめてから終了操作するまでを「1セッション」と言います。データベースの中で利用できる「検索結果の一時保存」「検索履歴の閲覧」などは通常そのセッションの中でのみ有効です。

よく使う検索式の保存などは、個人用のアカウントを登録して、そのアカウントの中で行います。こちらはセッションが切れても問題ありません。

また、多くの利用者にスムーズに利用してもらうため、セッションの途中で何の操作も行わない時間が続くと、システムがセッションを切ってしまうことがあります。検索の途中で、別の調査など始める場合は、重要な検索結果は保存しておき、こまめに終了するようにしてください。