

情報検索の過去・現在・未来：システム志向からユーザー志向へ

三輪眞木子

放送大学教養学部教授



【要旨】

最初に、サーチャーとしてクライアントのために情報検索をした場合と、自分が記事を書くために情報検索をした場合の情報検索プロセスの違いを紹介する。サーチャーとして検索する場合は明確に定義された情報ニーズが最後まで変わらないのに対して、著者として検索する場合は漠然とした情報ニーズが徐々に変化することに気付いたことを述べる。これは、システム志向の情報検索とユーザー志向の情報検索の違いである。

次に、情報検索への2つのアプローチである、システム志向の情報検索アプローチとユーザー志向の情報検索アプローチのモデルを紹介し、情報ニーズ、情報探索プロセス、適合性について両者の違いを述べる。また、新たにデータベースを構築する場合や、既存のデータベースをユーザー志向に再構成する場合には、徐々に発展する情報ニーズを支援し、ユーザーの主観的な適合性判定に配慮し、学習や調査のプロセスを支援する必要性を指摘する。

最後に、メタデータの重要性と、メタデータ標準に準拠することのメリットを紹介する。文字データを含まない、画像、音声、動画のコンテンツを検索する際には、メタデータが整備されていないと、コンテンツを見つけ出すことが不可能である。また、文字、画像、音声、動画のあらゆるメディアのコンテンツを収録するデータベースの相互乗り入れや、博物館、図書館、アーカイブズの連携に、メタデータ標準が果たす役割について述べる。

【略歴】

<学歴>

- 1973年 3月 日本女子大学文学部卒業
- 1978年 12月 Pittsburg 大学大学院修了 MLS (図書館・情報学修士) 取得
- 1983年 3月 慶應義塾大学大学院文学研究科図書館・情報学専攻博士課程単位取得退学
- 1994年 8月 Syracuse 大学大学院情報学科博士課程入学
- 2000年 5月 Ph.D. in Information Transfer 取得 (Syracuse 大学)

<職歴>

- 1980年 4月-1994年 9月 慶應義塾大学文学部図書館・情報学科兼任講師
- 1981年 6月-1983年 3月 筑波大学学術情報処理センター技官 (準研究員)
- 1983年 8月-1985年 3月 株式会社エポックリサーチに参加、取締役就任

三輪眞木子

1985年4月-1994年9月 株式会社エポックリサーチ代表取締役就任
1994年10月-2001年3月 株式会社エポックリサーチ取締役相談役
2001年4月-2009年3月 メディア教育開発センター教授
(この間、総合研究大学院大学文化科学研究科
メディア社会文化専攻教授等を併任)
2009年4月-現在 放送大学教養学部教授

【主要研究業績】

著書

『データベースと情報管理』放送大学教育振興会、2011年（共著）

『遠隔学習のためのパソコン活用』放送大学教育振興会、2012年（共著）

『情報行動：システム志向から利用者志向へ』勉誠出版、2012年（単著）

Quality Assurance in LIS Education: An International and Comparative Study, Springer, 2015年（共著）

論文

Makiko Miwa, Yuka Egusa, Hitomi Saito, Masao Takaku, Hitoshi Terai, Noriko Kando, „A Method to Capture Information Encountering Embedded in Exploratory Web Searches,”
Information Research, vol.16, no.3, 2011.

<http://www.informationr.net/ir/16-3/paper487.html>

三輪眞木子「検索のゆくえ」『情報の科学と技術』63(1)、2-8頁、2013年

<http://ci.nii.ac.jp/naid/110009562659>

三輪眞木子「海外における「図書館情報専門職」の質保障とコンピテンシ」『情報の科学と技術』66(2)、71-78頁、2014年

情報検索の過去・現在・未来：システム志向からユーザー志向へ

三輪眞木子
放送大学教養学部教授

はじめに

私は、1980年代から30年以上にわたって、プロフェッショナルとして、また、研究者として、情報検索に係ってきました。情報検索にコンピュータが導入された初期の時代には、記憶容量の制約のせいで、書誌データの検索、つまり、書籍の目録や論文の索引といった短い文字列の検索しかできませんでした。その後、論文の本文を電子化した電子ジャーナルが登場し、全文、つまりフルテキストを検索できるようになりました。1990年代にはインターネットが登場して瞬く間に世界中に広がりました。その間、コンピュータの記憶容量も急速に拡大して、文字データだけでなく、画像や音声や動画といったマルチメディアのデータベースが登場しました。今では、Web上で多様なメディアのデータベースを検索できるようになりました。そして最近では本を丸ごと電子化した電子書籍が普及し始めています。

本日は、情報検索の発展の歴史とこれからの展望について、私の経験談を振り返りつつ次のような点を見ていきたいと思えます。まず、私の情報検索専門職、通称サーチャーとしてのこれまでの経験の中で、システム志向の情報検索プロセスと、ユーザ志向の情報検索プロセスの違いを認識したきっかけについてお話しします。次に、情報検索の歴史における変遷を見ていきます。特に、文字情報の検索とマルチメディア、つまり、画像や音声や動画や立体のような文字を伴わない情報の検索の違いを見ていきます。そして、システム志向の情報検索の考え方と、ユーザ志向の情報検索の考え方の違いについて、具体例を交えてご紹介します。単一のデータベースシステムとして構築されてきた情報検索環境から、最近では、データベース同士の串刺し検索や相互乗り入れに移行している状況についてお話しします。つまり、博物館、図書館、アーカイブズの連携が進んでいる状況をご紹介します。そして、このようなデータベースの広がりの中で、情報を効率的・効果的に検索するための重要なカギとなるメタデータ、つまり情報の代替表現がなぜ重要かを考えていきます。

私の情報検索経験

それでは、まず、私の経験談から始めましょう。私は、大学卒業後に、外資系出版社で編集の補助のような仕事を担当していた時に、コンピュータ上で文字をコード化して扱うことができることを知り、データベースに興味を持ちました。日本では、コンピュータによる文字処理について教えてくれる大学院が見つからなかったので、アメリカのピッツバーグ大学の図書館情報学科の修士課程に入学して、文字の電子化とコンピュータ処理、つ

まりデータベースの構築と検索について学び、図書館情報学修士を取得しました。帰国後に、慶應義塾大学に出来たばかりの図書館情報学の博士課程に入学して、そこで日本語の書誌データベースの構築と検索に関する研究開発に係る機会を得ました。ただし、残念ながら、博士論文を書くことなく、満期退学しました。博士時代から、筑波大学の学術情報処理センター、今でいうところの基盤センターで、日本全国の大学などを対象に、学術データベースをオンラインで提供するサービスを担当していました。当時のボスは、「コンピュータは24時間動いているのだから、人間も同じように働くのが当たり前」という考えの持ち主で、昼間はデータベースのサービスや講習会の企画・運営などの業務を行い、夜は研究開発と論文の執筆に明け暮れる日々を過ごしました。

当時、アメリカでは、利用者に代わってオンラインデータベースの検索をする専門職としてのサーチャーが図書館や研究機関で活躍しており、情報ブローカーという情報の収集・分析をする会社が注目を集めていました。日本で情報ブローカーを起業しようと考えて、1983年にエポックリサーチという会社を設立しました。設立当時は、業務内容を理解していただくのが大変でした。資金繰りのため銀行に出向いても、前例がないということでなかなか融資を受けられませんでした。マスコミ等で知られるようになり、ようやく軌道に乗せることができました。この会社では、企業や公的機関の調査業務、データベースの設計、図書館目録の電子化、海外の同業者と連携して国際的な調査などを実施しました。業務内容は多岐にわたりましたが、いちばん成功したのは、情報検索のプロフェッショナルであるサーチャー養成講座でした。医療、特許、科学技術、ビジネスの多様なデータベースの検索を実習つきで実施して、その後に情報科学技術協会が実施している「サーチャー講座」の先駆けとなりました。また、個人としては、「サーチャーの時代」¹という本を丸善から出版して、専門書としてはベストセラーとなり、「サーチャー」という職業を社会に認知していただけるようになりました。一方で、慶應義塾大学をはじめとするいくつかの大学で、「情報検索」等の授業を非常勤で教える機会もありました。

1990年代になると、インターネットが登場して、情報検索の環境は変わりました。一番大きな違いは、それまで検索のプロフェッショナルであるサーチャーしか触ることのできなかったオンラインデータベースを、ユーザが自分で検索できるようになったことです。この変化を目の当たりにして、検索代行業の将来性に期待できなくなったため、慶應義塾大学で取得できなかった博士号を手に入れて研究者に転向したいと思うようになりました。そこで、博士論文の執筆を志して、1984年にシラキュース大学の情報学の博士課程に入学しました。当初は、3年間の計画でしたが、フルタイム学生として、研究や教育に係るのが楽しくて、結局6年弱の長期間にわたって、シラキュースで学生生活を送りました。シラキュースという町は、アメリカのニューヨーク州にあるのですが、カナダとの国境に近く、5大湖に近いので積雪が多いところで、11月から3月は雪に覆われた日々を過ごしました。研究に取り組むにはまたとない環境だったと思います。

2000年に博士号を取得して日本に帰国し、2001年に国立大学共同利用機関であったメディア教育開発センターに教授として採用されました。メディア教育開発センターは、その

後、独立行政法人となり、2009年には放送大学と合併、といえますか、吸収されて、現在は放送大学の教授をやっています。

エポックリサーチで、サーチャーとして仕事をしている際に、KDDが発行していた「On the Line」という月刊誌に、「サーチャーの冒険」という4ページの記事の執筆を依頼されました。これは、編集者と相談して毎回テーマを決めてデータベースを検索して、その結果をまとめるというもので、楽しみながら執筆することができました。この連載は、後日、「亜米利加解説新書」というタイトルで書籍として出版しました²。この連載をしているときに、サーチャーとして他者のために情報を検索する場合と、著者として自分の書く文章のために情報を検索する場合には大きな違いがあることが分かりました。

サーチャーとして情報を検索する場合は、最初にクライアントのインタビューによって情報ニーズと検索結果として何がほしいのかを明確にしたうえで、目的に合ったデータベースを選択して、適切な検索戦略を立ててから検索を実施します。想定した結果が得られない場合には、検索戦略を立て直して、納得のいく結果が得られるまで検索を続けます。ただし、使うデータベースは有料なので、短時間で効率的に検索を進めることが重要でした。クライアントの情報ニーズに見合った検索結果が得られると、それに目を通して、必要な情報を抽出して、解釈を加えてレポートにまとめてクライアントに届けます。その間、最長でも1週間くらいでした。時には、医療関係者から「大至急」という依頼を受けて、2～3時間で回答することもありました。

一方、著者として、記事を書くために情報を検索する際には、まず記事のトピックを決めるためにデータベース中の情報を探索しますが、その際の情報ニーズは漠然としていて、まずは仮のトピックを選定して編集者と相談します。次に、仮のトピックについて焦点を絞った検索をしながら、見えそうな記事や統計データを探し、得られた情報を解釈しながらストーリーを組み立てます。その際、うまくストーリーがまとまらなかったり、十分なデータが得られない場合は、トピックを変更して、編集者の了解を得ます。そして、組み立てたストーリーを一定の文字数に書き上げて、イラストレータに絵のイメージを伝えます。

これら2つの情報検索プロセスには、大きな違いがあります。サーチャーとしてクライアントのために情報を検索するときは、最初に定義した情報ニーズと想定されるアウトプットの要件は最後まで変わりません。また、検索結果が情報ニーズに適合しているかどうかは、抄録や要旨を見て客観的に判断しています。これに対して、著者として情報を検索するときには、最初に定義した情報ニーズ、すなわち、記事のトピックは、情報検索を進める過程で得られた知識によって徐々に変化していくので、最初のトピックは大抵変化します。また、検索結果を吟味する際には、本文を読んで記事の情報源として見えそうかどうかを主観的に判断します。

情報を検索しながら記事を創り出すという検索プロセスを、認知、感情、行動という3つの側面で段階的に見ていきましょう。最初の段階では、認知面はアイデアの生成、感情面は情報を歓迎するムード、行動面は、新聞記事や年鑑や業界誌記事を読み人と対話し

ます。第二段階では、認知的には広いトピックを定義し、感情面は冒険的で、行動は、トピック領域のデータベースを検索して最新の記事を探します。第三段階では、認知面はストーリーの組み立てで、感情面はフロー状態、つまり、熱中しているため他のことが入ってこなくなる状態になります。行動は、検索結果として得られた見えそうな文献を通読したり統計データを確認します。第四段階では、認知的には、データと情報を組み立てたストーリーに結び付けていきます。判断が伴うため、感情面では、批判的になります。行動は、ストーリーに適合したテキストやデータを抽出し選定します。第5段階では、認知的には統合で、感情は論理的、行動は、図表を作成して解釈を文字化します。最後の第6段階では、認知的には発表、感情的には想像、行動はストーリーの執筆です。

では、サーチャーとしてユーザの情報ニーズを仲介する場合の検索と、ユーザが自分で検索する場合の検索の違いを詳しく見てみましょう。サーチャーとしての検索を仲介型検索、ユーザ自身による検索を自己検索とします。まず、トピックですが、仲介型検索の場合はトピックは明確に定義されているのに対して、自己検索の場合は、トピックが漠然としています。検索プロセスは、仲介型検索では、系統だっていますが、自己検索の場合は系統だっていません。検索戦略は、仲介型検索では段階的な手順として構成されていますが、自己検索の場合は試行錯誤によるものとなります。感情は、仲介型検索では安定しているというかほとんど感情の動きがありません。自己検索の場合は、検索結果を見て一喜一憂する不安定なものとなります。知識構造は、仲介型検索では一定の方向に徐々に深まっていますが、自己検索では方向が変化して拡張していく感じです。そして、検索プロセスを貫くテーマは、仲介型検索では達成を目指していますが、自己検索では創造、つまり成果を生み出すことを目指しています。

この違いはまさしく、システム志向の情報検索とユーザ志向の情報検索の違いです。つまり、システム志向の情報検索では、情報ニーズは変化せず、適合性は客観的に評価されるのに対して、ユーザ志向の情報検索では、検索している最中に情報ニーズが変化し、情報の適合性は主観的に評価されるのです。この違いに気づいたときに、これで博士論文が書けそうだと思います。

情報検索への2つのアプローチ

情報検索システムの開発は、1950年代から進められてきました。インターネットが登場する前の1980年代までは、情報検索の専門家であるサーチャーが検索を担当するのが一般的で、システム志向の情報検索の考え方が基本でした。1990年代にインターネットが登場し、Web上のブラウザでユーザが自力で情報を検索するようになると、ユーザ志向の情報検索の考え方が登場し、現在は両者が拮抗しています。

システム志向の情報検索アプローチ

まず、システム志向の情報検索のモデル(図1)を紹介します。この図の左端にはデータベースに収録されている文献群があり、インデクサが各文献に索引つまりメタデータを付

与します。メタデータについては、後ほど詳しく説明しますが、ここでは、情報検索の際に文献の代替となるタイトル、著者名、出版社名、主題キーワードを指します。図の右側は、情報ニーズを持つユーザがおり、キーワードを組み合わせたクエリとして情報ニーズを表現してデータベースに問い合わせます。インデクサが文献に付与するキーワードと、ユーザがクエリに使うキーワードを対応させるための翻訳ツールとして、シソーラスや分類表や件名表目標が使われています。つまり、インデクサはこれらの翻訳ツールを参照して索引を付与し、ユーザはこれらの翻訳ツールを参照してキーワードを選定します。

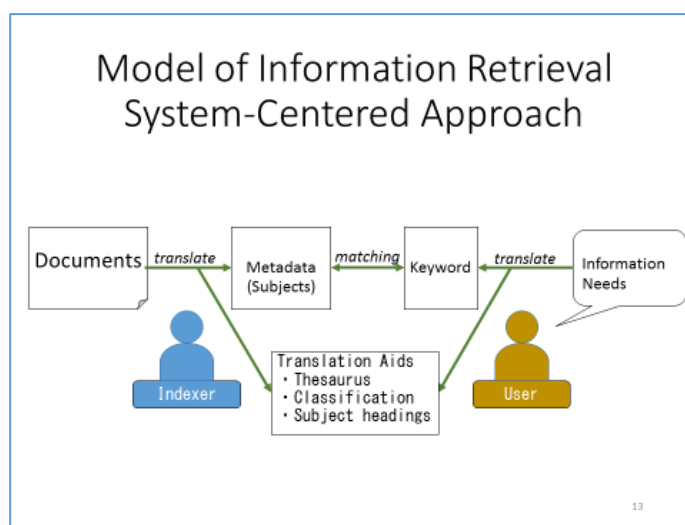


図 1：システム志向の情報検索モデル

ユーザは情報ニーズを、キーワード同士の論理演算でクエリに表現します。その際に利用できる論理演算は、論理和、論理積、論理差の3種類です。

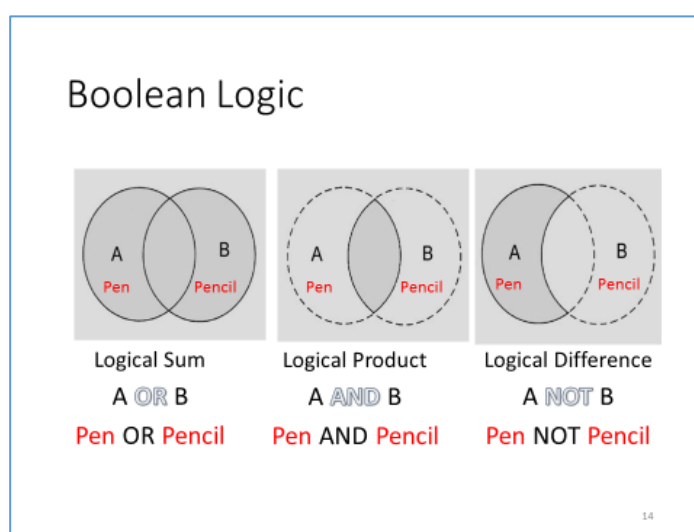


図 2：ブール演算

たとえば、「ペン」と「鉛筆」の2つのキーワードの論理演算を考えてみましょう。論理

和の場合は、「ペン」と「鉛筆」のどちらか一方もしくは両方を含む文献を検索します。論理積の場合は、「ペン」と「鉛筆」の両方のキーワードを含む文献を検索します。論理差の場合は、「ペン」というキーワードを含むが、「鉛筆」というキーワードは含まない文献を検索します。実際の検索では、論理和、論理積、論理差を組み合わせることで、複雑なクエリを表現できます。

情報検索では、最初に作成したクエリが情報ニーズに適合した検索結果を導くとは限りません。そこで、適合フィードバックという手法が考案されました。適合フィードバックは、最初の検索結果を修正することで、より良い検索結果を導く手法で、明示的フィードバック、暗示的フィードバック、擬似フィードバックという3種類の技法があります。

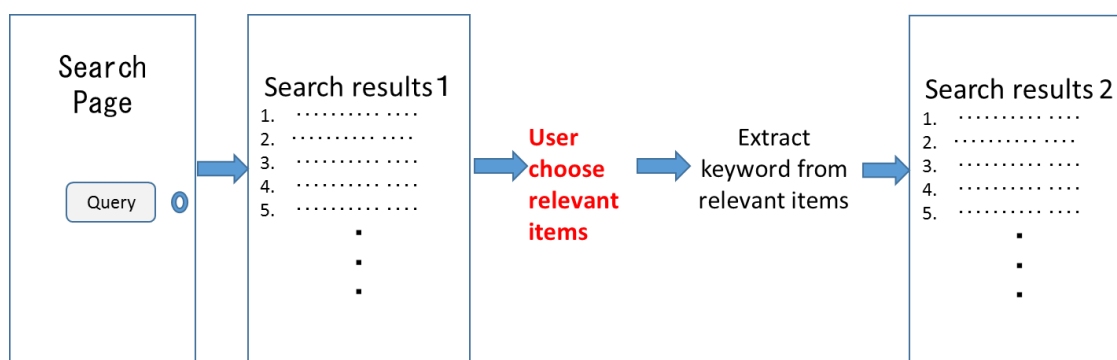


図3：明示的フィードバック

明示的フィードバック（図3）は、最初のクエリによって検索された文献の適合性をユーザが判定して、適合していると判定された文献のメタデータ情報を使って次のクエリを組み立てる技法です。

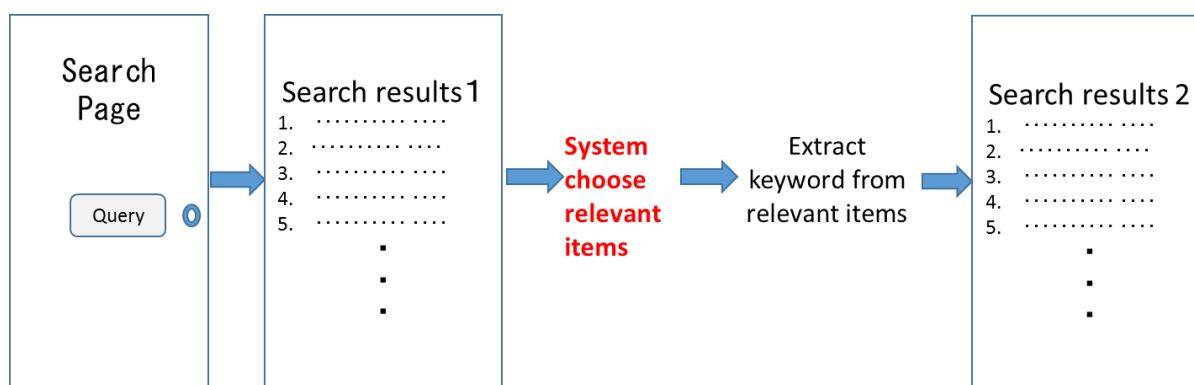


図4：暗示的フィードバック

暗示的フィードバック（図4）は、最初のクエリの検索結果中の各文献をユーザが見たかどうかや、見ていた時間や、ブラウジングやスローリング行動をシステムが感知して、それに基づいてシステムが適合性を判定します。適合していると判定された文献のメタデ

一タ情報を使って次のクエリを自動的に組み立てる技法です。ユーザが知ることなく適合性が推測されるため、暗示的フィードバックと呼ばれます。

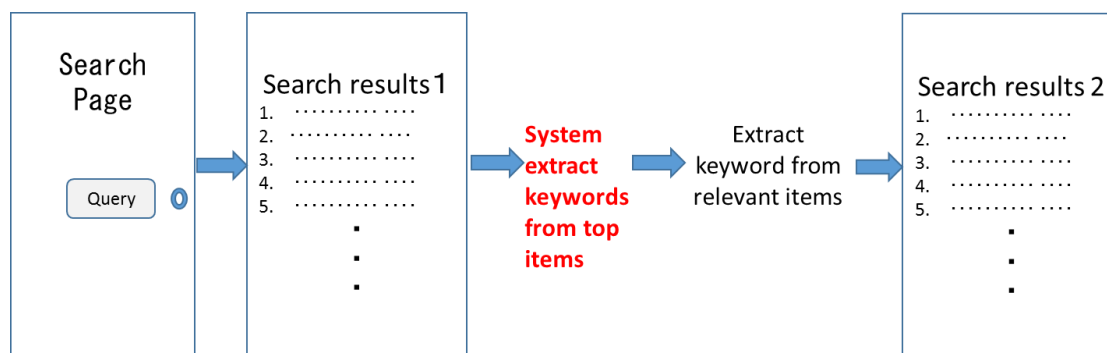


図5：擬似フィードバック

擬似フィードバックは、適合文献集合を検索した後、適合性の高い順に表示された検索結果一覧の上位数件を適合性が高い文献と判定して、それらのメタデータを使って、次のクエリを自動的に組み立てる技法です。

このように、適合フィードバックでは、適合文献の内容に基づいてクエリを修正することで、より適合性の高い検索結果を導き出すために使われています。システム志向の情報検索では、検索結果の適合性を高めるために、これらのほかにもキーワードの重みづけなどの様々な技法が使われています。

ユーザ志向の情報検索アプローチ

次に、ユーザ志向の情報検索アプローチの考え方を見ていきましょう。ユーザ志向のアプローチでは、情報ニーズを徐々に発展するものととらえています。図6は、情報ニーズの発展過程を示しています。

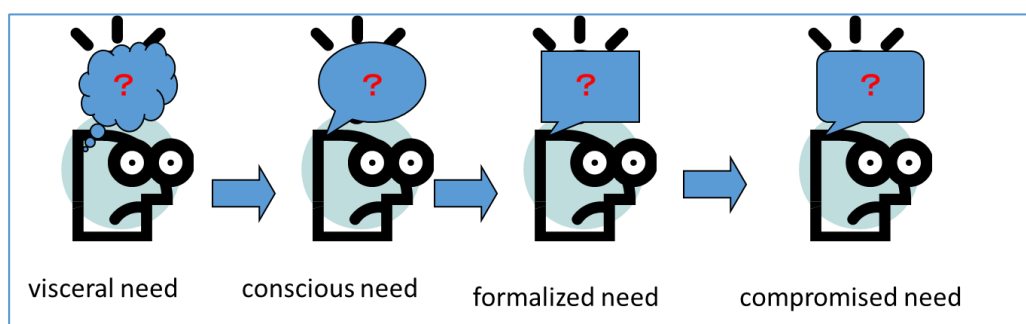


図6：発展する情報ニーズ³

この図に示す通り、情報ニーズは段階的に発展するもので、最初は、心奥のニーズ (Visceral need)、つまり、言葉では表現できない真のニーズです。つまり、漠然と違和感を感じては

いても、それが何かを自分自身がよくわかっていないので、他の人に説明できない状態です。次の段階は、意識されたニーズ（Conscious need）です。これは、十分には定義できない未決定の曖昧なニーズで、人との対話を通じて曖昧さを解消する可能性がある状態です。次の段階は、具体化したニーズ（Formalized need）で、自分の情報ニーズを明確に表現できる状態です。更に次の段階になると、妥協したニーズ（Compromised need）となります。これは、情報検索システムから回答を得られるように、システムの特徴に合わせてキーワードを組み合わせてクエリを組み立てたり、サーチャーが理解できるように情報ニーズを説明するという状態です。情報ニーズはこのように発展するため、初期段階ではユーザはうまく情報ニーズを表現するのが困難かもしれないというわけです。ユーザ志向の情報検索における「情報ニーズが変化し、明確に定義できない」という考え方は、「情報ニーズは明確に定義でき、変化しない」というシステム志向の情報検索モデルの前提と異なります。

次に、情報探索プロセスにおけるユーザの行動や認知の全体像を表現した情報行動モデルをご紹介します（図7）

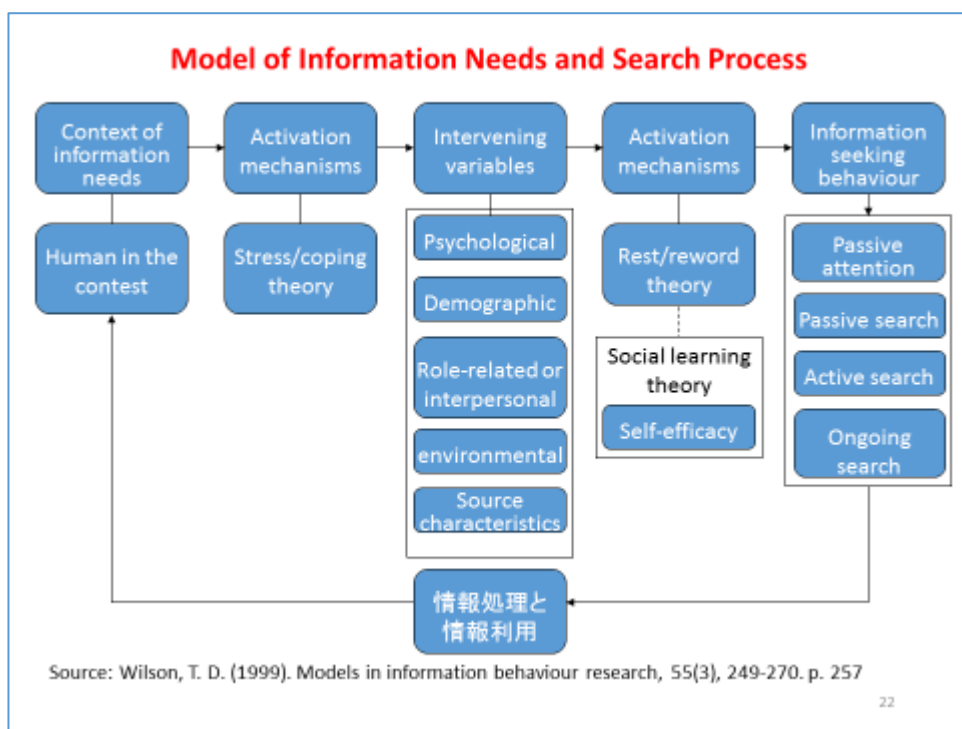


図7：情報行動モデル⁴

このモデルは、情報ニーズを持つユーザが、情報探索行動を起こす際に、ストレス対処理論に基づいて、情報を探す際のストレスを勘案して行動を起こすかどうかを決めることを示しています。情報行動を起こす際の介在変数として、ユーザの心理的な要因、社会経済的要因、役割や対人要因、環境要因、情報源の特性を挙げています。そして、情報探索行動の活性化メカニズムとして、リスク報酬理論に基づいて、情報を得ることによるメリ

ットと、情報を得ないことによるメリットを天秤にかけたり、社会的認知理論に基づいて、情報を探すことに対する自己効力感の程度によって情報検索をするかどうかを決めることを示しています。次の情報検索行動では、周囲の環境に目を配って異変がないことを常に確認する情報行動である受動的注意、世の中の出来事を把握するために新聞を読んだりニュースを聞いたりする受動的探索、変則的知識状態を解消して情報ニーズを満たすための能動的探索、特定テーマについて新しい情報を継続的に入手しようとする継続的探索という4種類の情報探索行動があることを示しています。最後に、獲得した情報を処理、つまり解釈することで、情報探索行動の一連のプロセスが完了します。

このように、ユーザ志向の情報検索アプローチでは、いくつかの重要な知見を提示しています。まず、知識が変則的であるために、ユーザは情報ニーズを明確に述べるのができないかもしれないということです。次に、情報ニーズを満たすような検索結果を提供するためには、情報ニーズの表現方法が重要であるということです。そして、ある種の情報ニーズは、情報探索行動を導かないことも明らかにしています。たとえば、体調がすぐれないのに医者に行きたがらない人や、深刻な病気だと診断されたのに自分の病気や最善の治療法について情報を求めようとはしない患者がいることが知られています。また、情報探索行動には、受動的なもの、能動的なもの、継続的なものがあることを明らかにしています。

情報検索では、ユーザの情報ニーズを検索結果が満たしているかどうかを、適合性 (relevance) によって判定します。システム志向の情報検索では、第三者が判定できる客観的な指標として適合性を位置づけており、それによって検索結果を判定しています。図8は、適合性を基準として検索性能を判定するための、精度 (precision) と再現率 (recall) の2つの尺度を説明しています。

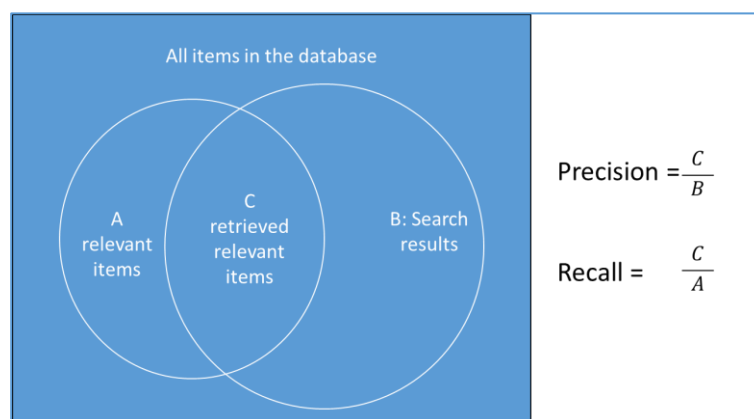


図8：精度と再現率

この図の青色の部分は、データベースに含まれている全文献を表しています。このデータベースに特定の情報要求に適合する文献が20件(A)あったとします。このデータベースで検索した結果、検索結果が40件(B)得られました。本来BにはAが全て含まれるべ

きですが、とり残しがあれば、それは「検索もれ」となり、A以外の無関係なものが混じっていれば、それは「雑音」となります。この「検索もれ」と「雑音」の尺度を決めたものが、再現率 (recall) と精度 (precision) です。検索結果のうち、適合する文献が 12 件 (C) であれば、再現率 $R=C/A=12/20=0.6$ (検索もれ=20-12=8) となり、精度 $P=C/B=12/40=0.3$ (雑音=40-12=28) となります。なお、ここで、R,Pともに1 (最大値) であれば、検索もれも雑音もない完璧な検索となります。

このモデルは、検索結果が適合しているかどうかを第三者が客観的に判定できることを前提としています。それに対して、ユーザ志向の情報検索の考え方では、適合性はユーザ自信にしか判定できないとしています。システム志向の情報検索では、検索結果の主題が適合しているかどうかを客観適合するので、適合性を測定できるのに対して、利用者志向の情報検索では、検索結果が文脈に適合しているかを主観的に判定するので、適合性の測定は懇談であるとみなしています。では、ユーザは検索結果を、どのように判定するのでしょうか？ まず、新奇性、つまり、ユーザがそれまで知らなかった情報を得られたかどうか、新しさ、つまり、情報が最新のものであるかどうか、品質、つまり、得られた情報が質の高いものであるかどうか、包括性、つまり、得られた情報に漏れがないかどうか、容易さ、つまり、ユーザが容易に読みこなせるかどうか、そして、アピール、つまり、ユーザの関心を惹くかどうかという観点から、多角的に評価しています。

情報探索プロセスにおいて、ユーザは何らかの目的をもって情報を求めています。そして、得られた情報を多面的に評価して、情報を獲得するとユーザ自身の知識構造が変化します。そのため、情報を求めているユーザの情報検索プロセスは、あらかじめ道筋を立てることが困難です。

ユーザ志向の情報検索の研究では、情報探索プロセスを描写するモデルが生み出されているので、それらを紹介しましょう。まず、ベリーピッキングモデルです (図 9)。

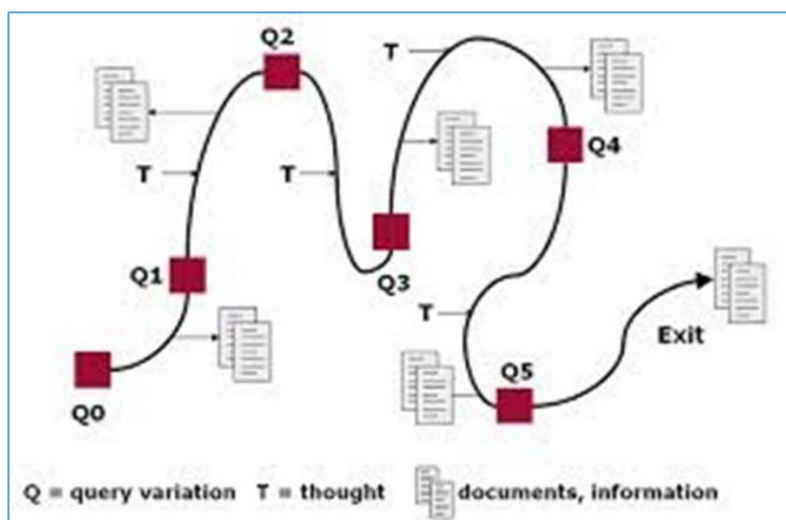


図 9 : ベリーピッキングモデル⁵

これは、野山で野生のぶどうを摘むやり方を情報探索プロセスに応用したモデルで、野

生のぶどうは灌木の間に分散していてまとまって生えているわけではないので、一度に一塊りずつしか摘めません。それと同じように、情報を探すユーザは、一つのまとまった検索結果としてではなく、少しずつ情報を収集します。また、文献データベースで提供されているものを越えた多様な探索技法を使っています。一般的に、次の6種類の検索技法が使われています。

脚注の追跡は、論文などの脚注や文末引用に示された文献を追跡する技法です。

引用の探索は、論文の被引用献をたどって新しい文献を見つけ出す技法です。

雑誌の通覧は、関心領域の雑誌の目次を定期的に通覧して文献を見つけ出す方法です。

領域スキャンは、関心対象領域で発表される論文や学会発表を通覧する技法です。

抄録・索引探索は、抄録誌や索引誌の通覧で、これがデータベースの検索です。

著者探索は、特定の著者が発表した論文や著書を探索する方法で、ホームページや SNS のプロフィールを探すものを含みます。

次は、初めてレポートを書く高校生の情報探索プロセスを描写した、情報探索プロセスモデルです（図 10）。

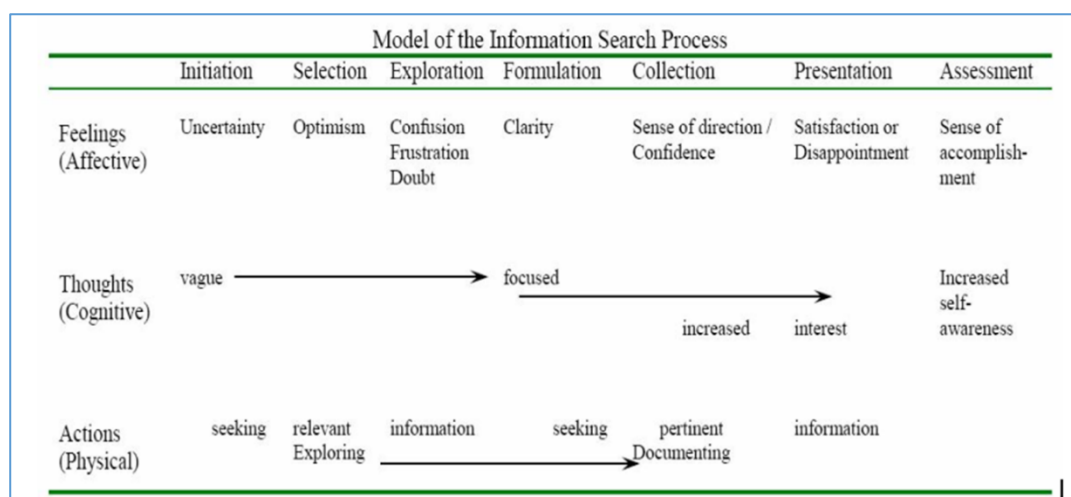


図 10：情報探索プロセスモデル⁶

第一段階では、教師からレポートの説明を受け、何をすべきかよく分からないため、不安になり、トピックの選択や取り組み方について人に相談します。第二段階では、レポートのトピックにどう取り組むかを決めますが、レポートの要件や、入手できる情報源、時間配分を想定して、トピックを選びます。トピックが決まると不安が消えて漠然とした希望が生まれます。第三段階では、選んだトピックに関する情報を収集してトピックへの理解を深めます。その際に、入手した情報が相互に矛盾するため、落胆と恐怖を感じてフラストレーションが高まります。第四段階では、知識構造が明確になるにつれてトピックの焦点が定まります。この段階では、トピックに関する文献を読み、テーマやアイデアを書き留めていく中で、不安が消えて自信が高まります。第五段階では、トピックの焦点に

直結した情報、つまり、レポートで使えるような情報を選別して、焦点の定まった詳細なメモを取ります。トピックの焦点への関心が高まるにつれて不安感がなくなり自信が高まります。最後の第六段階では、情報探索を完了し、見逃しがないか確認したうえで、執筆の準備に取り掛かります。トピックを個人的な観点で統合する際に、思考は最高潮に達して、思考がまとまり情報探索を完了すると、開放感を感じます。このモデルは、情報探索プロセス中に、思考や感情が変化すること、教師や仲介者は、生徒が情報探索のどの段階にいるかを想定し、適切な情報を提供する必要があることを示唆しています。

次は、研究者や経営者といった、情報収集にたけた人の情報探索プロセスを描写した、専門家の情報探索行動モデルです（図 11）。

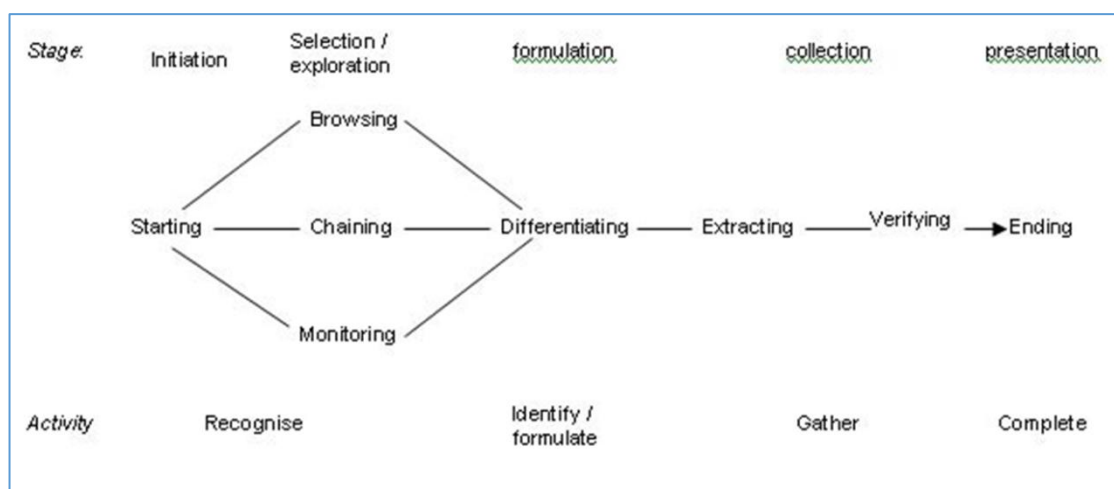


図 11：専門家の情報探索行動モデル⁷

情報を探し始める際には、そのことを知っていそうな同僚や知人に相談したり、図書館や Web で調べものに取り組みます。連鎖探索とは、入手した情報を手がかりに、他の情報を引き出す方法で、引用を使ったり、知り合いの紹介を受けたりするものです。ブラウジングとは、はっきりとした方向性を持たない、漠然とした情報探して、図書館や書店で関連分野の図書や雑誌を並べた棚をざっと見るのがその一例です。Web の関連カテゴリーをクリックして、どんな情報があるかをざっと見る場合もブラウジングにあたります。情報監視は、特定のテーマについての新しい情報が集まるようにするもので、知人や専門家にあるテーマの新しい情報が得られたら教えてほしいと頼んだり、データベースに特定テーマを登録しておき、そのテーマの新しい記事や論文が出たら自動的に知らせてもらうカレント・アウェアネスもその例です。次の情報源の選別は、情報源の性格に基づいて情報源を選別するもので、著名な雑誌に掲載された論文を選んだり、新聞に書評が掲載された書籍を選んだり、専門家が推薦する情報源を選ぶという行為がそれに当たります。情報抽出とは、情報源から有用な情報を抽出するもので、記事や論文や本を読みながらメモを取ることや、Web ページをブラウジングしながら重要部分をコピー&ペーストすることが該

当します。妥当性の検証は、情報の正確さや適切さを確認することで、複数の情報源から得た関連情報を照合することが該当します。もし獲得した情報に矛盾がある場合は、どちらが正しいかを追求します。新しい情報が得られなくなった一求める情報が見つからないとか、時間切れによって、探索を終了します。

これらのモデルは、ユーザの情報探索プロセスの次のような特徴を示しています。

- ・ 情報探索プロセスはループや試行錯誤を含む線形のプロセスです。
- ・ 最初は大きなトピックから開始して、徐々に狭いトピックに絞り込んでいきます。
- ・ 情報探索プロセス中に、情報ニーズは発展していきます。
- ・ ユーザは自分の判断を継続的に評価しています。
- ・ 情報探索プロセス中に知識構造は変化します。
- ・ ユーザは、情報探索プロセスを終了する何らかの方略をもっています。

このような、ユーザ志向の情報探索は、Web 上の情報探索プロセスモデルに反映されています (図 12)。

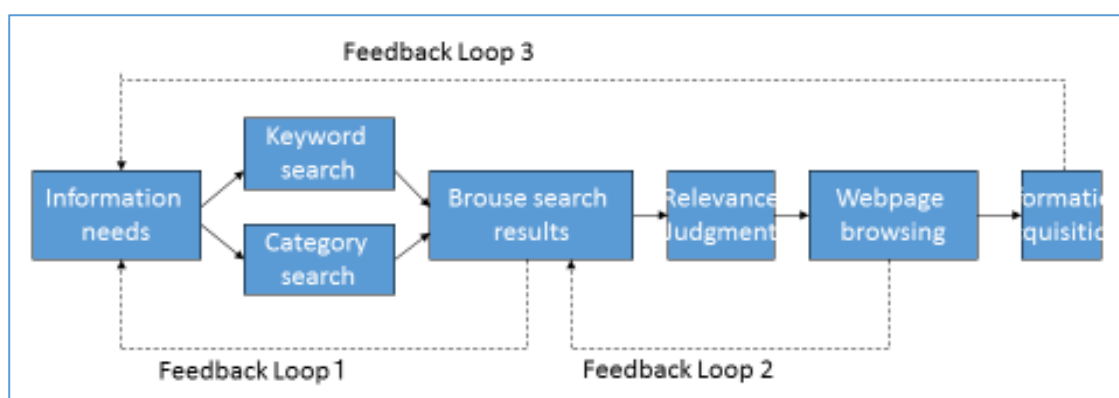


図 12 : Web 検索プロセス

ユーザは、情報ニーズを感じると、Web で検索します。検索方法は、情報ニーズが明確であればキーワード検索を、漠然としているとカテゴリ検索を使うと思います。検索結果のリストをブラウジングして、情報ニーズに適合しているものをクリックします。適合しているものが見当たらないと、情報ニーズを見直して、キーワードやカテゴリやサーチエンジンを変更します。表示された Web ページをブラウジングして、求める情報を獲得します。

求める情報が得られないと、検索結果のリストに戻り、情報ニーズに適合した別のものをクリックします。求める情報が得られないと、情報ニーズ自体を見直します。そして、ベリーピッキング・モデルに示したように、獲得した情報から新たな疑問、つまり情報ニーズが生まれて、次の情報探索が始まります。

ユーザ志向の情報検索では、探索型検索という概念が使われています (図 13)。従来型の情報検索とユーザ志向の情報検索を対比したこの図は、情報検索を、参照 (Lookup)、

学習（Learn）、調査（Investigate）の3種類に分類して、そのうちの学習と調査のための検索を探索型検索と位置付けています。ちなみに、参照型検索は、システム志向の情報検索で、現在の情報検索システムは、参照のための検索を支援しているけれど、学習や調査のための検索は十分には支援していないと考えられています。ですから、これから新たにデータベースを構築したり、既存のデータベースをユーザ志向のものに刷新する際には、学習と調査の支援を考慮する必要があります。

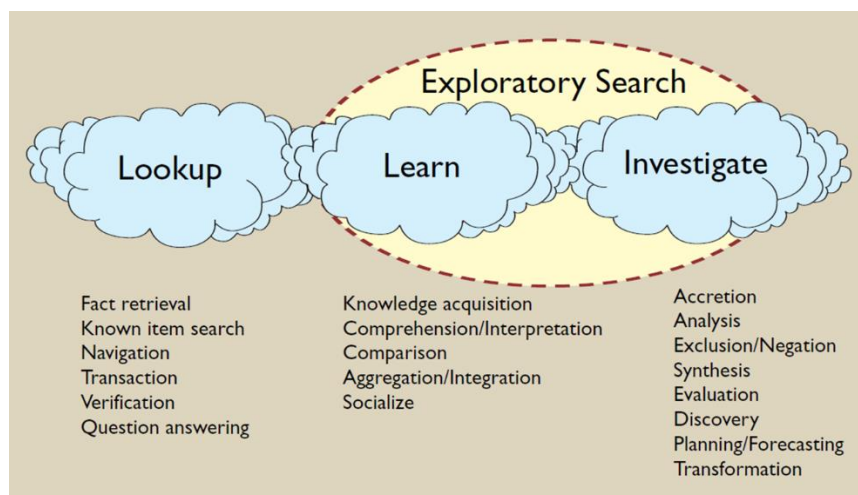


図 13：探索型検索⁸

メタデータの重要性

情報検索、特に主題検索における重要な要素はメタデータです。メタデータとは、「データのデータ」という意味で、情報検索では、コンテンツを代替するものと位置付けられています。これは、インデクサが付与するキーワードのことです。

メタデータの役割を視覚的に表現したイメージを見てみましょう（図 14）。



図 14：メタデータのイメージ

このように、メタデータは情報の中身が何かを示すラベルです。電子オブジェクトを検

索する際に、コンテンツが文字データであれば、文字によって検索できますが、文字を含まない写真や音声や動画の場合は、メタデータが整備されていないと、検索できません。

メタデータの主な目的は、適合情報の発見を支援すること、情報管理のために情報源を組織化すること、及び、ユニークな ID を提供することです。伝統的なメタデータの例に、図書館の目録、名詞、ディレクトリーがあります。電子オブジェクトのメタデータは、特定領域のメタデータ標準を使って作成することで、情報検索だけでなく、相互乗り入れなどを支援して、オブジェクトの有用性を高めることができます。電子オブジェクトのメタデータ例には、Web ページのメタタグに記述するタイトルや言語や制作ツールなどがあります。Wikipedia の記事カテゴリー、タイトル、情報源、アクセス方法も、メタデータです。

表 1 は、各ジャンルの主なメタデータ標準です。

表 1：各ジャンルのメタデータ標準

genre	Metadata Standards
Library	Dublin Core, MARC (Machine Readable Catalog)
Digital government	e-GMF (e-Government Metadata Framework)
Museum	CIDOC CRM
Geography	FGDC (The Federated Geographic Data Committee)
Education	IEEE LOM (IEEE Learning Object Metadata)
Rights	Indecs, XrML (eXtensible rights Metadata)
Multimedia	MPEG-7, MPEG-21
News articles	NewsML
Broadcasting	TVAIF, SMPTE, P/META, Jmeta

図書館では、ダブリンコアがメタデータ標準として使われています。ダブリンコアは、電子政府、つまり、アーカイブのメタデータ標準である e-GMF や、電子教材のメタデータ標準である IEEE LOM の基本要素としても使われています。

ダブリンコアのメタデータスキーマは、表 2 に示す 15 要素で構成されています。タイトル (Title) は、オブジェクトの名称、制作者 (Creator) は、オブジェクト制作の主たる責任者で、主題 (Subject) は、主題を表すキーワードです。概要 (Description) は、オブジェクトの概要を説明する文章で、出版者 (Publisher) は、オブジェクトを入手できるようにする責任者、貢献者 (Contributor) は、オブジェクトの制作に貢献した制作者以外の関係者です。日付 (Date) は、オブジェクトのライフサイクルで生じた主な出来事の日付で、制作日、公開日、改編日などを記述します。タイプ (Type) は、オブジェクトの性格を表す要素で、雑誌論文、写真、統計データなどを記述します。フォーマット (Format) は、オブジェクトの物理的な形態で、Mime タイプで記述します。識別子 (Identifier) は、オブジェクトのユニ

ークな ID で、URI を使うか、URI がない場合は URL で代替します。情報源 (Source) は、オブジェクトの元になった文献等です。言語 (Language) は、オブジェクトに文字や声が含まれている場合の言語を記述します。関係 (Relation) は、オブジェクト同士の間関係や類縁関係を記述します。範囲 (Coverage) は、オブジェクトの位置や時期を表すもので、写真であれば撮影場所を、歴史的なモノであれば、年代を記述します。最後の権利 (Rights) は、著作権に関する情報を記述します。メタデータの記述は、文字を含まない画像や音声や動画の場合には、特に重要ですが、同時に、手間のかかる作業です。ただし、メタデータが記述されていないと、見つけ出すことは不可能なので、なるべく詳細なメタデータをつけることが必要不可欠です。また、メタデータを付与する際には、領域やジャンルのメタデータ標準に従うことも重要です。

表 2 : ダブリンコアのメタデータ要素⁹

ELEMENT	DEFINITION
Title	A name given to the resource.
Creator	An entity primarily responsible for making the resource.
Subject	The topic of the resource.
Description	An account of the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Type	The nature or genre of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A related resource from which the described resource is derived.
Language	A language of the resource.
Relation	A related resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Rights	Information about rights held in and over the resource.

メタデータがなぜ重要かについては、既に十分説明したと思いますが、一つは、異なるデータベースの串刺し検索 (Federated search) を可能にするためです。つまり、一つのクエリで複数のデータベースを検索する際に、クエリが複数のサーチエンジンに受け渡されて、全体から得た検索結果がまとめて表示されます。その際に、メタデータが違っていると、検索はうまくいきません。最近では、博物館、図書館、アーカイブズの連携が推進されていますが、相互のコンテンツを連携するには、共通のメタデータスキーマを使う必要があ

ります。同じジャンルのデータベースの国際連携をする場合も同様です。また、最近では、地域の博物館、図書館、アーカイブズ、自治体などが独自に地域資料の電子化を進めているようですが、メタデータが標準化されていないと、ユーザは全部まとめて検索することができなくなってしまう。

まとめ

最初に、サーチャーとしてクライアントのために情報検索をした場合と、自分が記事を書くために情報検索をした場合の情報検索プロセスの違いを紹介しました。その中で、サーチャーとして検索する場合は明確に定義された情報ニーズが最後まで変わらないのに対して、著者として検索する場合は漠然とした情報ニーズが徐々に変化することに気付いたことを述べました。これは、システム志向の情報検索とユーザ志向の情報検索の違いです。

次に、情報検索への2つのアプローチである、システム志向の情報検索アプローチとユーザ志向の情報検索アプローチのモデルを紹介して、情報ニーズ、情報探索プロセス、適合性についての両者の違いを述べました。そして、これからデータベースを構築する場合や、既存のデータベースをユーザ志向に再構成する場合には、徐々に発展する情報ニーズを支援し、ユーザの主観的な適合性判定に配慮し、学習や調査のプロセスを支援する必要性を指摘しました。

最後に、メタデータの重要性と、メタデータ標準に準拠することのメリットを紹介しました。特に、文字データを含まない、画像、音声、動画のコンテンツを検索する際には、メタデータが整備されていないと、コンテンツを見つけ出すことが不可能であることを指摘しました。また、文字、画像、音声、動画のあらゆるメディアのコンテンツを収録するデータベースの相互乗り入れや、博物館、図書館、アーカイブズの連携に、メタデータ標準が果たす役割について述べました。これらの知見を、アーカイブズやマルチメディアデータベースの効率的・効果的な検索の実現に活用してください。

1 三輪眞木子. (1986). サーチャーの時代：高度データベース検索. 丸善, 224p. (三輪眞木子. (1992). サーチャーの時代：高度データベース検索 (第2版), 丸善, 240p.

2 三輪眞木子・一ノ渡勝彦. (1991). 亜米利加解説新書. 世界文化社, 248p.

3 Taylor, R.S. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, vol.29, no.3, p.178-194. (図は筆者のオリジナル)

4 Wilson, T.D. (1999). Models in information behavior research. *Journal of Documentation*, vol.5, no.3, pp.249-279. (図は、p.34のFigure C3.3)

5 Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, vol.13, no. 5, p.410, Figure 2.

6 Kuhlthau, C.C. (1994). Seeking meaning: A process approach to library and information services, Norwood, NJ. Ablex Publishing, p.82.

7 Ibid. 4, p.160, Figure 25.1.

8 Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communication of ACM*, vol. 49, no. 4, Figure 1.

⁹ Dublin Core Metadata Initiative. (1995). Dublin Core Metadata Element Set, Version 1.1: Reference Description. <http://dublincore.org/documents/1999/07/02/dces/>