

投稿論文

「情報検索」と知識社会

奥 雅 雄

1. 何のために探すのか

その場的なことから、生涯を掛けたものまで探し求めるものは多様である。

(1) 生産と消費

生存のために食料を求めるもっとも基本的な活動である。生物が生きていくためには獲物を探して捕食を行わなければならない。効率的に捕食を行える種は繁栄する。一方で自身が獲物として食べられる運命にもある。食物連鎖のなかに組み込まれて生存競争のなかで種を残す。

人間の場合、何段階も画期的な変化が起きて食餌行動が変わってしまった。

大脳の進化し、知的能力が発達した

話し言葉を獲得して協働作業ができるようになった

「もの」や「ものごと」をシンボル化して理解できるようになった

文字を発明し、時空の制約を少なくすることができた

農業を行い、生産性が飛躍的に高まった

道具・機械を使うようになった

情報処理の仕組みを発達させた

食物の生産が飛躍的に進み、多くの時間を生産活動以外に使えるようになった。結果として、食物以外のものも豊になり、文化が発展して、多様な方法とスタイルにこだわるようになった。それゆえ、生まれてから学習する内容が増え、一人前になるまでの期間が大幅に伸びた。蓄積された知識を探すのにあるいは学習に時間がかかるようになった。

(2) 子孫を残す

もう1つ求めて探す必要があるのは種の保存機能としての異性との出会いである。人口密度がもっと低い時代では再生産のための枠組みが必要になる。食べ物獲得の競争のなかで、種族間の争いがある一方で、略奪もあるので必ずしも友好的とは言

えないかもしれないが姻戚関係があった。個人的な意思よりも種族間のおきてが強かった。

知・知識は一族のなかで子どもたちに伝えられ、子どものなかでも年長の子どもが年少者をリードする方式でも教育が行われた。こうして一族の繁栄が保証されるようになった。

出会い系サイトはあまりよい響きをもっていないが、企業間の関係では有効に活用されている。顧客として、共同事業者として、問題の解決者としてパートナーが捜し求められる。1回きりの取引も多いが、最初の出会ってから永続的な関係にも発展する。場合によると地域の産業を興し、経済の活性化につながる場合もある。かつての「家」は現代の「企業」として永続が期待されている。

(3) 知識の獲得

言葉は文字よりはるか以前からあったから、言葉によって多くの知識を獲得した。縄文時代でも季節による獲物の違いや冬が来てそれに備えるような生活様式ができていたと予想される。正確な暦の存在はわからないとしても季節に相当するような概念を持っていたと思われる。ある時期になると、豊かな魚介類あるいは、秋の実りは多くのものが得られたし、そのことは理解していたはずである。

栽培農耕が始まって、やがて稲作が始まる。さらに多くの知識が必要とされ、収穫を確保するためには暦は相当に重要なものになる。病虫害や天災との戦いの歴史であつたろう。奈良・平安時代に米の品種が記録されているので想像より豊かな知識を持っていたと考えられる。その他、水利のための土木工事、政治のための戸籍や条里制など多くのものを大陸から学んでいる。

政治や荘園など管理の必要性から文書の重要度が増し、役人は書記官としての意味をもち、当然、日本では1300年前から、中国ではその倍程度前から文章力が必須の能力になっていた。文書があることは、書きっぱなしではなく、後から取り出して、読み返すことも必要になり、図書館のような書庫が存在した。当然、何らかの分類法や索引などが利用されて過去の資料を探すことが行われたと予想される。ほとんどは記憶の範囲で十分であつたであろう。

文字の歴史は官僚と同時に詩人を生み、学者を育てた。官僚がそれを目指さないと文字を必要としないから、歴史のはじまりでは、詩人の多くは官僚でもある。李白、王維、大伴家持、藤原定家など。

(4) 知識の利用

一番の目的は「もの」そのものを得ることである。現代では購買行動となって何らかのものを得ることになる。

社会のサービス化にともなって、ものの概念が広がり、実体のないものが多くなった。結婚式場探し、切符・チケットの購入、CD・DVD からダウンロードによるコンテンツの購入、保険・貯蓄・投資、教育・学習など、有形のものを含めて拡張してオブジェクトと呼ぶ。

オブジェクトを獲得するためにそれらの所在や入手方法をメタ情報と呼ぶ。直接、オブジェクトを得る行動をする前に情報を得る行動をした後、本行動をとることになる。オブジェクトがこだわりのものほど厳選されるようにメタ情報が重要になる。A社の製品とB社の製品がどちらでも良いのでなく、差別化された違いが重要なのである。ものが豊富になって、生活者のこだわりが増し、オブジェクトの選択に時間がかかるようになった。奇しくもネットが提供する検索機能はこれらの情報を得るのに時間的にもコスト的にも時代にマッチしたものになっている。もの作りに情報が重視され、社内では、取引記録から情報を作り出し、社外では動向等の情報に目が向けられる。マーケティング活動によって直接顧客行動についての情報を集め、商品計画がおこなわれる。

一方、消費行動の場合も知識や情報にかかわることが多くなった。小説、ドラマ、映画、テーマパークなど知的で頭脳的・精神的な消費が多くなっている。

これらの生産・消費活動はすべて知識・情報とのかかわりが大きい。

2. 検索の背景

(1) 検索の歴史

検索は、蓄積されたデータに対して行われる。情報システムが登場する以前は、図書カード、図書目録、辞書・辞典、本の目次・索引などが利用されていた。一方で検索と対極にあるのは自然そのものから観測・観察、実験によって対象を解明する方法である。社会現象なども自然に含めて考える。直接得られたものを1次情報とすると、検索結果は2次、3次のもので、標準化等の形式美が整えられるが、鮮度の点で注意が必要になる。

(2) 整理整頓と「もの」の取り出し

探す方法の基本は整理整頓である。この方式は分類に基づいて整理・保管する方法になる。家庭内で「もの」が行方不明になる最大の原因は整理・保管ができていないことである。保管場所もビジネス現場と較べて体系的でない。引き出し、押し入れ、納屋など種種雑多な保管場所があり、ものがあふれるに従って古いものが下へ下へと隠れていく。

IC カードなどでも使われている IC チップは IC タグとして「もの」へ貼り付けて利用する。バーコードの次世代の仕組みとして期待されているが、家庭内のもの

へ貼り付けておくと簡単に探しだせるようなシステムが期待できる。実用的かどうかは別として、現代の生活はものばかりがあふれていることが最大の問題となっている。

分類はカテゴリーという。商業でいうと業種に該当する。化粧品店、本屋さんの分類である。現代では多くの商品を扱い業種で分類しにくくなっている。あらゆるものを扱い、売り方に特徴を持たせた業態が重視される時代である。コンビニ、総合スーパー、ドラッグストア、ホームセンターなど若干の扱い商品の違いがあるが、何屋さんとはいえない。調べたいことも分類では迷うことが多くなる。境界にまたがる事象をどう扱うかが課題になり、人によって連想が異なれば見つけられなくなる。

(3) 検索の誤り

膨大な情報が蓄積した現在の検索ではキーワードによる方法が主流である。

検索には過誤が伴う。統計でも登場する第1種の誤りと第2種の誤りで次の違いがある。

①第1種の誤り

第1種の誤りは、正しいのに正しくないと判断されることで、検索結果に出てこない。そのため情報として拾いようがない。その他、次のようにいわれることもある。

一般的にはあわて者の誤り

生産者危険ともいわれ、過剰品質になりやすい

②第2種の誤り

正しくないのに正しいと判断される。検索では不必要な結果が数多く出てくる。100万件、1000万件と出てきても意味がない。その他、次のようにいわれることもある。

一般的にはばんやり者の誤りとされる。

消費者危険ともいわれ、欠陥品や不良品を受け入れるはめになる。

この「第2種の誤り」を少なくしてより適切な結果を返す仕組みが次世代のウェブとして期待されているセマンティックウェブである。

(4) 蓄積と活用

文明・文化が進んで膨大な知識の蓄積は、簡単には習得できなくなった。人生の

大きな期間を学習にあてることになる。専門化して知識を担う人々の領域が細分化している。キーワードが思い浮かばないときは、直接、誰かに聞いてみるのが好都合である。対話形式のブログは検索がうまく行えないときに便利さを発揮する。対話のなかでは未熟者の意を汲み取って丁寧に答えてくれる。

自分の気ままな検索は、一回きりの場合やしつかりした方向性をもっている場合は有効に働くが、方向性や体系的なコントロールは期待できない。対話的に軌道修正が効くと回り道をしなくて効率的な学習効果が得られる。それゆえにインストラクター、トレーナー、教員との直接的な接触に意味が出てくる。補助的に検索の仕組みを利用する場合や目的意識を持った検索が有効になる。

3. 検 索

(1) 文書中の検索

Word 中の文言を探すには検索機能を利用する。単なる文字だけでなく、書式検索、ワイルドカード、あいまい検索などが可能になっている。

文字を扱うソフトでは少なくとも文字の検索・置換機能は持っている。

同意語は検索の対象になりにくい。「酒」、「日本酒」、「清酒」は微妙の差異があり、検索から漏れる場合がある。「バッグ」、「鞆」、「かばん」も意識して指定しないと漏れる。「バイオリン」と「ヴァイオリン」は字面は異なるが Word の「あいまい検索」では拾い出すことができる。これらの差異・近接関係を記載する辞書が「シソーラス」で、類語辞典、類義語辞典ともいわれる。検索機能に関連語を含めることもできるが、使うソフトによって違いが生じる。Word では半角・全角、かな・カナの違い、「バ」、「ヴァ」等の表記の違いに対応している。

(2) 保存資料の検索

Windows では、ファイルやフォルダを探す機能がある。ファイル内の文字は WindowsXP の場合すべてではないのでサービスパックをインストールして、未登録の拡張子のファイルにインデックスを付ける機能を選択する必要がある。

これらは従来、テキストファイルに対する grep 機能であったが、使われている語句のインデックスを作って検索するので最初は時間がかかるがそれ以後は早く結果が出てくる。各種のバイナリーファイルの場合も内容をチェックできるものが多い。

ブラウザへインストールするソフトとして、次のツールバーが各検索サイトから提供されている。自分のパソコン内の検索ができる。

Yahoo! Desktop Search (YDS)

Google デスクトップ検索 (GDS)

Windows Desktop Search (WDS)

Vista へは Windows Search (XP でも利用可能)

検索時に気をつけることは文字コードの差異である。日本語は次のいくつかの方式があり、検索で指定する文字と一致しないことがある。通常はソフトが自動的に判断する。

ユニコード (この中でもいくつかの方式がある)

EUC

S-JIS (マイクロソフトの当初の方式)

JIS (汎用コンピュータで利用されている)

(3) ネット上の検索

ネットでのサイト検索は次の3者がある。

Google

Yahoo

MSN

いずれもアメリカの企業である。日本では Yahoo が利用されるが、世界的には圧倒的に Google が強い。

検索の仕組みは次の2段階で行われる。ユーザーの要求に対して世界中に探しに行くとしたら、応答まで何日かを要す。即座に結果が出てくることはあり得ない。

① ロボット検索

無人のコンピュータが世界中のサイトを定期的に訪れてインデックスを作成する。各ページに何が書かれているかを要約し、データベースへ保存する。

② ユーザーへの応答

キーワードが体系的に整理・保管されたデータベースとなっているので即座に該当情報を送り返すことができる。

(4) 文字以外の検索

画像や映像のライブラリーが作られ蓄積物が増大している。これらはシンボル化しないと検索は難しい。文字の題名や説明などを検索の対象にする。

音楽などでは、メロディなどから検索可能であるが、曲名などの文字情報が検索の主たる方法になる。

地図は地名からポイント検索の他、直接、ポイントをクリックして詳細化したり、

隣接地を選ぶ方法で取り出せる。

地図も画像の一種として、画像や映像、音楽や音響の自動的な意味づけは難しい。同じ見ためでも歴史的な意味などは別に説明を加えないと分類や検索は困難になる。絵や音に歴史・文化が重層的に重なっている。

(5) セマンティックウェブ

あまりに雑然とした検索結果に苛立ちや不満を感じる人が多い。賢い検索システムとして、今、研究が進んでいる。

①タクソノミー

再び分類が出てくるが、キーワードをどのような大系で考えるかが課題でタクソノミーという。シソーラスという用語辞書を作り出す作業になる。

たとえば、「モデル」にはグラビア誌などの人としてのモデル、モデル店舗、モデル校、モデル住宅など「もの」や「仕組み」に対してのモデル、モデル実験、数学モデルなど研究や開発で使われる「模型」や「実験設備」などを指すものまで多様にある。

人の場合はグラビアモデルとか、ファッションモデルとして検索すると厳密に探せる。それらの意味情報をもたせてより明確な検索を目指す方法がオントロジーと呼ばれている。とくに RDF: Resource Description Framework では単純な検索で用いられるキーワードに変わって、メタデータとして主語、述語、目的語に分けて記述する方式をとっている。これによって意味を正確に表現できるので不必要な検索結果を表示しなくてよくなる。

②フォークソノミー

タクソノミーは絶対基準として図書館の書籍分類のような体系化が重視されるが、はたしてそのような単一の価値基準で分類できるだろうか。

この点を考慮したものがフォークソノミーといわれるもので、立場や視点を異にした分類基準が数多く出てくることを許す。

独特の世界で生きている人々に外部から規定できるような基準は使えない。独特なのだから独特の関連性が生じ、当然、当事者グループに理解できる構造ができあがる。今、タクソノミーの限界に対して、フォークソノミーの可能性が広がりつつある。具体的にはタグを利用して自分たちだけの分類を作ることが可能な世界である。それだけ検索しやすい仕組みを作っていけることになる。従来の検索がトップダウンで押しつけ的に作られていたのに対し、人々の草の根的にボトムアップに検索の仕組みを構成できることである。

4. 知的活動の開花にむけて

先人の知識を汲み取ることも必要だが、その先にある未踏の世界を探索することも価値がある。研究機関の研究だけでなく、ベストヒットを生み出すような工夫も宝物の探索となる。

単純な検索一行動ではなく、幾重にも重なった知識を獲得して、ものごとが成し遂げられる。一連の知識の創造活動こそが新しい時代と新しい価値を創り出す。

膨大な知識の中で習得して使いこなすために知的大系はますます複雑になっている。思想と道具の有効な活用が不可欠になっている。

(1) 知的集積と知識社会

ビジネスでは構成メンバーがビジネス情報を共有し、品質の高い商品・サービスをタイムリーに提供する仕組みが必要である。

研究段階では産官学の連携が重視され、新しい時代の技術や社会的位置づけなどが研究されている。産官学の「産」とは産業界の意味で、個々の企業を指す。「官」は官公庁の意味で国や地方自治体の施策をいう。最後の「学」は研究機関としての大学などの研究機能を指し、一緒になって力を合わせて未来を切り開く姿勢を呼んでいる。商品やサービスのレベルが上がり、技術などが高度化しているので用いられる知識も膨大になる。すでにエレクトロニクスだけでの開発では不足し、メカ（機械）やソフトなどの融合が重要になっている。マーケティングからは、デザインや使い心地（ユーザーの声）などが重視される。

これらの知的な蓄積が企業の活力を生み、絶え間ない新商品の提供が企業の存続を約束する。知的蓄積（知識）は、商品の開発に優位さをもたらし、マーケティングでの顧客のこころをしっかりととらえることができる。

これらは一言で表現しているが、実際にはデータベースとネットワークに裏付けられた情報システムがバックボーンを構成し、分析力や構想力をもったスタッフによって維持されている。「知」は人間の活動に大きく依存している。

個人の知識は消費生活を豊かにする上で重要である。インターネットによって、消費行動が様変わりしている。

単純にネットによる通信販売が時間的な支援になり、広範囲な入手が可能となって便利さと選択の幅の広がりを与えただけではない。

ものの選択に大きくかわかり、消費者自身がネットに情報を提供するようになった。CGM: Consumer Generated Media といわれる仕組みで、ウェブに消費者を取り込んで活用する方式をいう。アマゾン.comでは本に対する評価を消費者に委ねている。☆マークや意見をページに取り込んでじっくり判断して、本が買えるように

している。消費行動を分析して、おすすめ本を羅列する「この本を買った人はこの本も買っています」などとしている。

消費財についての知識は当初、まわりの人の意見から、雑誌などの評価へ移ったが、店員からのアドバイスなどで買っていた。いまでは大半の情報がウェブやブログなどから得られている。購入寸前の消費者ははるかに大量の知識をもって商品を選ぶことが多くなった。

ブログ、wiki、ユーチューブなどもユーザーのパワーを結集したものになっている。

(2) データベース

知識の蓄積が図書館からデータベースに移行しつつある。図書館の役割は変わらないし、知の位置づけにおいても図書館は重要な役割を担っている。

調べものをするときの安易さが手伝って、図書館や書籍が置き去りにされる傾向にある。

辞書は、紙から電子辞書に移り、一般用語などは Wikipedia が利用される。漢字の書き方などは携帯電話の変換機能が利用されている。

電子辞書は紙の辞書に較べて効率的になったが、インターネットの利用方法は問題である。紙の書籍は情報源として保証がされているが、インターネットでは種類雑多なものが入り交じっている。即時的に鵜呑みにするのは安易すぎる。多数の意見を集めるなどの裏付けが必要になる。あるいは、信憑性を判断するだけの見識が求められる。関連情報などからその分野に精通し基礎的な判断力があれば有効に活用できる。

大きな問題を表面的に検索して理解したつもりになることは、使い方としてよろしくない。ほんとうに貴重な情報は簡単には手に入らない。

(3) ネットワーク

知・知識を取得する場所が、ネットへシフトしている。直接的体験の視点から危惧する意見も多い。ネットに限らず現代の生活は擬似的な仮想体験が大部分を構成する。ネット以前はテレビが問題視された。それ以前は本や新聞が仮想的な体験を提供していたし、今も本・新聞やテレビがかなり利用されている。それらの擬似的体験が無意味とは思えない。ただネットに偏りすぎることは検討する必要がある。

不足がちな実体験をこれらのメディアが補い、より充実した実践が行えるなら、メリットが期待できる。あくまでもバランスがとれているかの問題である。

大災害など、あらゆる先端的な機器類が使えなくなる状況で、サバイバルを実現するためには、人間本来の能力がものをいう。平生の日常生活で、ときには先端機

器を使わないで、自然に近い方法をとることは意味があるといえる。野外活動や農業・園芸や日曜大工的に自分でものを作り出すような体験や人々と直接に触れるような体験が求められている。

(4) 知識の量と質

いまから50年も昔であれば、そんなに多くの知識を必要としなかった。たとえば、何かの店を開くとしてもその業種の狭い範囲の知識があれば開店できた。蕎麦店などではそば打ちの修行が必要なのでだれでもが開くことはできなかったが、修行中に掟のようなものを学んで年季があげればのれん分けが期待できた。

現代では思考内容が複雑化し、情報量が桁違いに多くなった。ICTの発達にともなって情報爆発とでもいえる状況になっている。一方でインターネットなどによって、仕入先などを探すことはできる。業界を知らなくてもそれらは簡単に集められ、担当者と接触することでより専門的な知識を得ていくことができる。この点が検索の本質になっている。

知識は人が担い、人を求めて遭遇することで貴重な知識や情報が得られる。

知識は組織や集団で共有され活きて進化発展している。最先端のものごとはそれらの組織・集団に関わっていないと手に入れることはできない。それらはインターネットなどで公表されることもないし、公表されるような内容はすでに古いものになっている。

研究機関、企業の開発部門、プロジェクトなどに最先端の知識が蓄積され、メンバーの経験をとおして次へつながっていく。

企業に属し、その企業の商品やサービスを考え、消費者を満足させることは企業の構成員として成長できる。1つのことを5年、10年続けると中堅になり、ベテランへ近づく。顧客の考えることを先駆けて考え、検索にヒットするような商品・サービス・機能を提供することが優良な社員の手本である。一人でするのでなくチームで進めるのでメンバーになればだれでもできる。メンバーに加わるのが重要で、それが成長の機会であり、「知」の共有の姿である。うまくいった組織が有用な「知」を増やし、「知識社会」を推し進めている。