



『情報探索入門』2002より

薬学研究科 金子周司

『京都大学に在籍する1名の教官について、主な研究テーマ、所属学会、学位論文、受賞歴、科研費の取得状況、最近の学会発表、最近の著作や論文、代表的な論文の被引用状況と当該論文掲載誌のインパクトファクターを調査せよ』

あたかも自己点検評価のような上記の調査は、附属図書館提供の全学共通科目『情報探索入門』のある1コマにおいて、学部2回生以上からなる受講生に課している演習課題である。受講生は思い思いの教官について、メディアセンター端末から国立情報学研究所NACSIS-IR、米国ISI社Web of Science、あるいはインターネット上に点在するWebページを渡り歩きながら情報を収集し、レポートを作成して電子メールで提出する。

この演習は、あらゆる学术データベースがインターネット技術と融合し、電子図書館の整備が行き届いた京都大学においては、様々な学术情報をインターネットで学内から入手できつつあることを、受講生に体得してもらうことを目的としている。しかし、同時にインターネットが万能の百科事典ではなく、図書館や研究室に



累積する印刷物による情報と相補的な関係にあることに気づいてもらう意図もあり、何人かの学生は演習レポートでその点を鋭く指摘してくる。

『情報探索入門』は電子図書館を含む図書館の諸機能を系統的に学ばせる講義・演習として平成10(1998)年度からスタートした。教授陣として長尾眞総長、佐々木丞平附属図書館長、川崎良孝教育学研究科教授、カールベッカー総合人間学部教授、佐藤理史情報学研究科助教授と筆者の6名に加えて、演習補助者として図書館職員有志15名にもご尽力いただいている。

筆者はこの中でインターネットやデータベース、いわゆる電子図書館について担当してきたが、わずか5年間でインターネットを介して流通する学術情報の量と質は格段の進歩を遂げてきた。本稿では、実際の講義と演習で取り上げている題材から、2002年の現状を紹介したい。

講義

筆者は分子薬理学の研究者であるが、生命科学の最先端では「情報革命」とでも呼ぶべき研究情報の流通に関する変化が起きている。身近な例は電子メールの普及であり、知人との日常的な連絡はもちろん、いまや顔も知らない海外の研究者とメールを介して知り合い、共同研究を行うという話は特に不思議なことではなくなった。

この講義では、生命科学領域でWebを利用した3つの事例を取り上げて紹介した。

1. HighWire Pressプロジェクトに見られる論文審査・公表形態の変化

1995年、米国スタンフォード大学図書館に端を発した電子出版HighWire Pressは、論文の電子出版を強力に推進してきたプロジェクトである。J. Biol. Chem.(JBC)誌の全論文をAdobe社のportable document format(PDF)形式でインターネットに「出版」して以来、多くの学会と出版社

を包含し、現在では医学抄録データベースMEDLINEに収録される4500余の学術誌のうち332誌のオンライン化を果たし、さらに加盟数は増加している。JBC誌は1997年より機関購読契約制度を開始しており、当時、薬学部がJBCの機関購読をわずか年間16万円の追加投資で契約したところ、京大キャンパスの大部分からPDF論文が発行されたその日に読めるようになった。これが現在、数千もの学術誌を研究室に居ながらにして読める電子図書館時代の始まりである。

電子出版は現在、学術誌の電子編集へと進化したつつある。電子編集においてもJBC誌はWebとメールを駆使した優れたプロトコルを採用している。一般的な学術論文の審査過程においては、論文投稿者、編集者、査読者(レフェリー)の間を印刷された原稿やレフェリーコメントが何度も渡り歩く(図1左)。例えば、京大の研究者が米国の学術誌に投稿し、その査読が英国とフランスの研究者に回されて、その間の通信がすべて郵送によるならば、通信に費やされる時間は延べ十週間にも及ぶことになる。JBCの電子投稿・編集システムでは、論文原稿を執筆者がPDF化する(一部のワープロと画像ファイルも認められる)ことによって、Webページへの原稿PDFのアップロードが投稿に取って代わ

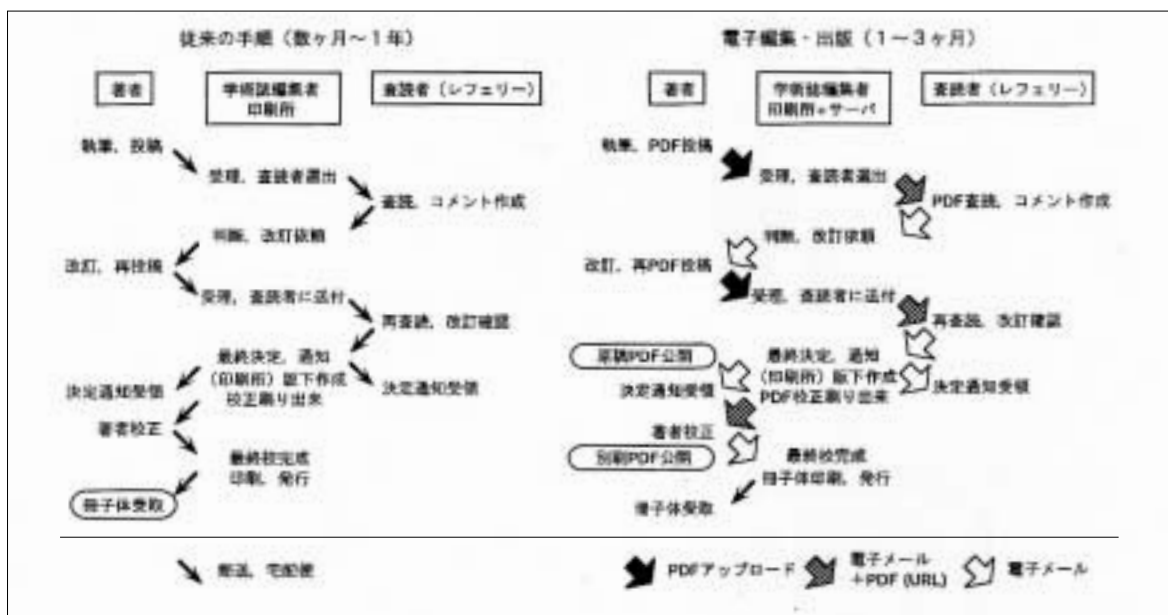


図1 論文投稿から学術誌の発行までの手順 (1回の改訂を要した代表例)

っている(図1右)。執筆者、編集者、査読者間の連絡はすべて電子メールによって行われ、隠されたWebサーバ上に置いた投稿原稿のURLを互いに連絡することによって、インターネット経由で原稿を見ながら審査が進んでゆく。採択が決定されると、ただちに論文原稿はそのままの体裁で仮出版され、公開されると同時に抄録データベースMEDLINEにも収録される。

この方式を採用した結果、JBC誌において最も迅速に審査が行われた場合、投稿からわずか2週間で原稿PDFが公開され、その後、2ヶ月以内に冊子体と同じ体裁の別刷PDFが取得できた。従来、数ヶ月を要した論文審査期間は電子編集によって画期的に短縮されることになったのである。

なお、JBCの場合、電子出版を採用した1995~2001年の間に年間ページ数が58%(約18000ページ)も増加した一方、インパクトファクターは1997年から2001年にかけて最低6.963(1997)から最高7.666(1999)の範囲でほぼ一定を保っている。この論文数の伸びは、同様に発展が著しく、かつインパクトファクターにも大きな変動のない神経科学領域のJ. Neurosci.が同期間に18%、総合領域のProc. Natl. Acad. Sci. USAが23%の伸びである事実と比べても突出している。すなわち、JBCが早期に電子出版、

電子編集へ移行した決断は、高い影響力を量的に示す学術誌としての成功を導いたと言えるだろう。今後は学術誌の生き残りをかけて、このような電子出版・電子編集が国内でも広く普及すると考えられる。

2. Entrez - PubMedに見られる研究データベースの統合化

ヒトゲノム情報に代表される遺伝子情報は原則的に公開されているだけでなく、タンパク質のアミノ酸配列や立体構造の情報、ゲノム上の変異や疾患、さらにはそれらを記載している文献抄録情報と統合的にリンクしたデータベースが形成されている(図2)。

米国でいち早くこれを可能にした背景として、遺伝情報センターNational Center for Biotechnology Information(NCBI)がNIHの医学図書館National Library of Medicine(NLM)の中に設置されている機構形態を指摘したい。すなわち、図書館が本来有している情報収集とデータベース機能が、遺伝子情報を蓄積、運用するにあたって最も優れた場所と見なされ、分子生物学の進展に伴って指数関数的に発生した遺伝子・ゲノム情報がデータベース化されていった。一方、かつてはDIALOGなどの検索ホストなどを通じて有料で使用するものだった医学文献抄録MEDLINEが1998年よりPubMedとして



図2 Entrez - PubMedホームページ (http://www.ncbi.nlm.nih.gov/entrez/)

無料でWeb公開されることになり、これは遺伝子情報との相互リンクを有するEntrezとして統合された。現在では、先に述べた学術誌の電子出版に対するリンクも設けられており、研究者は、例えばある遺伝病に関連した原著論文の別刷から、その遺伝子・ゲノム情報の詳細までを、すべて机上で調べられるようになった。

このようなインターネット上において無料で使用できる研究資源は、実験による実証を主体とする生命科学の研究手法に大きな影響を与え始めている。すなわち、従来は人間の観察力や経験、限られた情報から立てられた仮説を元にして実験を行っていたが、最近ではコンピュータによって抽象化された一定の規則性に基づく推測が、実験仮説あるいは考察として重要な意味を持つようになってきた。これは例えば、遺伝子断片の配列から抽出される相同性に基づく遺伝子の分類や機能の推定、一定の機能を有するアミノ酸配列（モチーフ）に基づく機能の類推などがあり、バイオインフォマティクス Bioinformaticsと呼ばれている。より将来には、普遍的法則の抽出、細胞や代謝のシミュレーションなどへの発展が期待される分野である。

この例は、情報の蓄積が新しい学問領域を芽生えさせた例としても見なすことができる。図書館も含めた情報学が今世紀の学問に及ぼす影

響力の大きさを予期させる事例である。

3. ライフサイエンス辞書プロジェクトでの研究資源の活用と公開

前2者に比べて極めて些細な例であるが、筆者が関わる電子辞書プロジェクトは一研究者が起こした研究成果の公開例としては当事者も予想外の経緯をたどっている。

事の始まりは1993年、パソコンで専門用語の漢字変換が不十分なことに業を煮やした筆者が、学会や出版社に専門用語集の電子化公開の予定について問い合わせたことにある。結局、生命科学のような学際領域の専門用語は、既存の辞書に頼らずに自作するしかないことがわかり、当時パソコン通信というメディアを通じて全国の有志を募ったところ、全員が異なる大学・学部から成る7名が集まり、科研費の助成を得られたことから辞書制作を始めた。用語はMEDLINE抄録の英文をコンピュータで計量的に解析するという手法で収集し、「多数決の論理」によってよく使われる用語から順番に収録した。当初から電子化して公開することを目的としていたため、少しずつ改良した辞書をパソコン通信で配布していたが、その頃、現在のWeb技術が完成したことから公開媒体をホームページに移行させた（図3）。



図3 ライフサイエンス辞書ホームページ (<http://lsd.pharm.kyoto-u.ac.jp/>)

現在ではオンライン対訳辞書をはじめ、用例、共起表現（ある単語の前後に出現する単語を示すことで頻用される語法が分かる）、音声など、書籍では不可能あるいは収録数に限界のある大量の情報から自由に検索できるページを公開している。また、携帯電話での検索や逐語訳を応用した学習教材の提示システムなど、新しい試みも行っている。

この電子辞書のホームページは、一日2万回以上の辞書検索をKUINS経由で世界中から受けているが、そのサーバは少し性能が良いだけのパソコンである。結局、こういった草の根的な情報発信は、その情報が唯一あるいは独自であれば、インターネット上では検索ロボットでヒットすることになり、価値あるものとして共有されるようである。最近では多くの研究者が自らのホームページを立ち上げて独自に研究情報の発信を行っているが、それが意味あるものかどうかは、専門に関連したキーワードを検索エンジンに入力して評価してみるとよいのだろう。

このオンライン辞書の利用者は、意外なことに、当初、想定していた学生や初学者よりも製薬企業やプロの翻訳者のほうが多いようである。制作者としては辞書を常にアップデートしなくてはならないという使命感を感じているが、この仕事に使える時間との葛藤に陥っているのが実情である。もしこの拙文を読んでくださって興味を持たれた奇々な先生がいらしたら、ぜひWebサイトのオンライン辞書を試みていただき、ご意見やご批判を頂戴できれば幸いである。

演習

『情報探索入門』では講義以上に、演習に力を注いでいる。筆者の担当はインターネット演習とデータベース演習の2回であり、前者では、演習補助者である図書館職員の方々が考案された10種類ほどの課題から、学生が興味をもった2つの課題についてインターネットの検索エン

ジンやディレクトリサービスを駆使して情報収集するというものである。課題は例えば、京都市の年間平均降水量を調べよ、といった簡単なものから、韓国に留学するための制度、手続き、費用を調べよ、という壮大な人生計画を練り上げるものまで多岐にわたっている。時にはこちらも意地悪く、インターネットでは決して正解にたどり着けない課題を出したりするが、実は図書館で教科書を開けば必ず載っている常識だったりしたこともある。そして、もう1回分の演習課題が冒頭に記した人物調査である。

いずれの演習も、あらゆるインターネットあるいはデータベース資源を調べ上げることを要求するが、ネット内の探索を繰り返すうちに、学生はインターネットにはちょっと昔の古い情報が少ないとか、収録されている情報の種類や分野に偏りがあることや、情報の信頼性が疑わしいケースに気づいてくる。つまり、学生は情報の信頼性にまで踏み込んで自分が得た情報を吟味することを体験する。このようにして、結局は電子メディアと従来のメディアを的確に使い分けられる学生を育てたいとの思いが、この演習には込められている。ともあれ、学部の2回生が冒頭に列挙したような事柄を、内容の理解はともかくも、情報としてはほぼ的確に収集できるほどに京都大学のインフラは整備されてきている。

インターネットを介した情報流通が学術研究あるいは大学において中心的なメディアになりつつあることが、現実味を帯びてきた。そんな2002年における『情報探索入門』の一部分を紹介した。

なお、講義で示したPowerPointレジュメ、演習課題およびレポート講評については、2003年5月まで下記で公開している（その後は2003年度の内容に更新）。

<http://www.media.kyoto-u.ac.jp/edu/lec/skaneko/>
(かねこ しゅうじ)