

第89回
(2015.12.1、12.8)

『インターネット検索の落とし穴とデータベースの活用方法』
金子周司教授(薬学研究科)

●第8回講義(12月1日)...受講生数:16名

・配布物:スライド

http://www.kulib.kyoto-u.ac.jp/modules/support/content/2015tansaku/INDB_20151201.pdf

◇講義目的 <インターネット情報およびデータベースとその活用法>

・インターネットやデータベースの現状を理解し、得られる情報の効用や特性について述べる事ができる。

◇講義内容

○学術情報とインターネット

- ・研究と文献(literature): 研究の価値は「originality(世界初)」。研究には「1.研究テーマを決める、2.文献情報から仮説を立てる、3.実験や調査をする、4.結果を文献と合わせて考察する、5.論文を書いて公表する」といったプロセスがあるが、2.のプロセスでまず何が既に研究されているかを漏らさず文献調査することが大事。
- ・「論文」の一般的構造: 1.表題と抄録(後に二次情報として検索される情報となる)、2.序論、3.実験方法、4.結果、5.考察、6.引用文献。このフォーマットを用いることで、他人による検証が可能となる。
- ・最先端の研究に必要なもの: その学問領域の基本的な知識、研究対象についての最新の知識、研究や調査をする技術、研究コミュニティ(人脈)、語学力、プレゼン能力。情報は発信元に集まる。
- ・文献=学術情報?: 学術情報は図書や雑誌といった永続性が保証されている文献のほかにも、データベースやウェブページのように書き換え可能で、必ずしも永続性が保証されていない情報もある。学術情報の形が変わってきている。
- ・学術情報は本来、インターネット向き: 世界共通、早さ、少部数、論文数が膨大、検索できないと意味がない、引用等がリンクできると便利、等の特性から、学術情報はいち早く電子化されるようになった。
- ・学術論文が出版されるまで: Peer-Reviewによる内容保証。採択率は1~3割程度が多い。また編集作業も電子的に行われることがほとんど。
- ・冊子体のないOpen Access ジャーナル: 従来の情報を得るものが買う購読モデルから、発信者が負担するOpen Accessへ。PLOS.orgでは読者の動向がリアルタイムに表示される。
- ・CCライセンス: 従来、著作権は出版社にあり、作者にはないことが多かった。しかし、現在作者が決められる著作権の意思表示の取り組みがある。(creative commons ライセンス) PLOS.orgなど一部はCC表示で再利用を広く認めている。
- ・電子書籍(Kindle/iPad): 電子書籍の普及が進んでいる。電子書籍では全文検索やマーキングも可能であるし、1台のデバイスの中に大量の書籍を入れておくことができる。
- ・データベースとは: データの集合体。データの検索と再利用が目的。索引づけ(インデクス作成)により高速な検索が可能になった。
- ・文献検索: 従来は書誌情報DBの二次情報から冊子の一次情報にあたっていたが、現在はダイレクトに電子的な形で一次情報にアクセスできるようになった。

●PubMed

- : 医学図書館 NLM の下部組織として遺伝情報センターNCBIを設置
- : 研究者が作った遺伝子情報と図書の書誌情報が融合
- : 今では化合物や疾患情報も統合されている
- : 統制語 MeSH が柔軟な検索を可能に

○ライフサイエンス辞書 Life Science Dictionary (LSD)

- ・なぜ電子辞書を作ったか: 卒業研究や英語での論文執筆や国際学会の発表など英語で苦勞する学生・大学院生が多い。需要はあっても大変なので電子用語集は誰もつくりたがらない。
- ・オンライン辞書サービス WebLSD: 英和・和英・出現頻度・発音・用例・共起検索・PubMed 連携
- ・どうやって作ったか
 - : 1. Native speaker の PubMed 抄録などから大量のテキストコーパスを作成 (fair use に該当する)、2. 単語の出現頻度や隣接数を数えて用語を収集(英語、日本語それぞれに)、3. 同じ意味をもつ語句を対訳とする、4. MeSH ツリーと照合する(上位語下位語を整理)。
 - : 専門用語は数多くの同義語を持っているため、表記のゆれが大きい。統制語を用いて制御する。

:関連性の算出(suggest 機能)、シソーラスとしても用いることができる。

- 応用例** マウスオーバー辞書(複合語検索も可能)、Kindle 辞書と PDF 辞書、iPad コンテンツ(無料の英単語帳アプリ)、電子カルテからの情報抽出。
- LSD プロジェクト** コーパスに基づく専門用語の分析、論文執筆を助ける共起検索、専門英語教育への活用、シソーラスからテキストマイニングへ。

○演習に向けて(ネットの落とし穴)

•**ネット上の学術情報** ネットの情報の網羅性は必ずしも高くない。背後のシソーラスは不明である。

•ネット検索での落とし穴

:流言・風説(鵜呑みにしない。客観的・科学的な証拠を探す。)

:ランク変動(二次情報の更新スピードが速い、持続性が低い。オプション指定で期間を限定する。キーワードの選択。)

:リンク切れ(有料データベースを調べる。キャッシュを調べる。)

:日本語と英語(必要なら英語でも調べる)

:伝達の範囲(英語で全て探せるかというところでもない。)

:同姓同名(場所や内容で判断)

•情報検索のポイント

:情報源(一次情報)までさかのぼる

:日本語の情報だけで満足しない

:いつの誰の情報か確かめる

:重要な情報は何か考える

:ネットだけに頼らない(書籍、専門家)

:真実や正解があるとは限らない

■**データベース演習** 課題の概要、目標の紹介。代表的データベースを使えるようになること、検索した情報の要約ができること、情報が無い場合はその原因を考えられることが目標。

•**データベースの紹介** 京都大学教育研究活動データベース、KAKEN、CiNii Articles、J-GLOBAL、Web of Science

•**論文と学術誌のステイタス**:他者に読まれて引用(Citation)されることが価値を決める(Impact Factor という指標)。研究者が投稿誌を決める際に利用される。Journal of Citation Reports にて調べられる。

•**新聞記事をデータベースで探す**

•研究業績検索のポイント

:複数の情報源をあたり、比較する

:本当にその人かどうか確かめる(同姓同名も多い)

:いつの情報か、発信者は誰か確かめる

:重要な情報は何か考えて整理する

:情報が無い場合の理由を考える

:ネット検索もしてみる

●**第9回演習**(12月8日) … 受講者数:17名

◇**演習ページ URL**:http://www.kulib.kyoto-u.ac.jp/modules/support/index.php?content_id=39

■**課題** 下記の京大に在籍する教員の中から1人を選び、下記データベースを利用して研究業績を調査せよ。

○**目的**:データベースの仕組みを理解して、効率的な情報収集を行うことができる。

○**注意点**:同姓同名の人物を見分ける。何が重要な情報かを考える。

■調査レポート項目

1. 氏名、現職

2. 研究開始年がもっとも新しい科学研究費補助金の獲得状況

3. 2005年以降の論文で最も被引用数の多いものの書誌情報(日本語1件、英語1件、計2件)およびその被引用数。共著者がいる場合はすべて解答すること。

※被引用数が不明または0の場合は、出版年月日が最も新しい論文を答えること。

※Web of Scienceの被引用数を解答する際は、Core Collectionからの被引用数を解答せよ。

- ・4. 教員に関する2005年以降の新聞記事
- ・5. 検索結果を考察し、以下の観点で論じよ。
 - (1) 同性同名の人物はどのデータベースで混じっていたか、またどのようにして調査対象人物を絞り込むことができたか。
 - (2) それぞれのデータベースの特徴を比較してまとめよ。
 - (3) 論文データベースと新聞データベースで得られる情報に違いはあったか。違いがある場合はなぜか。
 - (4) その教員の研究テーマの中で、最新の研究テーマ、周囲から評価されているテーマを挙げよ。また社会的にどのような点で評価されているか述べよ。

■使用するデータベース

- ・京都大学教育研究活動データベース
- ・KAKEN (国立情報学研究所)
- ・CiNii Articles (国立情報学研究所)
- ・J-GLOBAL (科学技術振興機構)
- ・Web of Science Core Collection (トムソン・ロイター)
- ・聞蔵IIビジュアル(朝日新聞)
- ・毎索(毎日新聞)
- ・ヨミダス歴史館(読売新聞)

■京大に在籍する教員 (この中から1人を選んで調査せよ)

- ・水野一晴 (文学研究科)
- ・鈴木基史 (法学研究科)
- ・文世一 (経済学研究科)
- ・林達也 (人間・環境学研究科)
- ・太田耕司 (理学研究科)
- ・今井裕 (農学研究科)
- ・荒井修亮 (フィールド科学教育研究センター)
- ・林春男 (防災研究所)
- ・中川一 (防災研究所)

◇補助内容: データベース選定、スライド資料(演習のポイント)作成、課題作成、演習ページ作成、演習補助、DB講習会の案内、採点

演習全体を通しての昨年度よりの変更点・感想・反省点など

◇昨年度よりの変更点

- ・演習が1回にひき、データベースの利用により重点を置いた演習を行った。
- ・例題をなくし、演習のポイントの説明を10分程度行ってから、課題の解答を開始した。
- ・課題の調査レポート項目に、新聞DBでの新聞記事の検索を加えた。
- ・調査対象の教員について、自由選択から多肢選択式に変更した。
(日本語論文、英語論文および新聞記事全てがヒットする教員を選んだ。)
- ・検索結果の考察について、考察内容が深まるよう、1つ1つの観点について、項目を立てて問うこととした。
- ・PandAの課題ツールを利用し、メール添付ではなくPandA上での提出とした。

◇感想・反省点

- ・課題提出者は16名、平均点12.81点(15点満点)、得点率85.4%。
- ・課題の難易度はちょうどよかったのではないかな。
- ・科研費の解答で、単位を間違えたり総額ではなく単年度の金額を答えたりするケアレスミスが目立った。解答時間中にアナウンスしてもよいかもしれない。
- ・学生のアンケートで、「正答がある問題はできれば答えを知りたいです。」という声があった。考察問題以外の項目については、解答を後日公開してもよいのではないかな。
- ・英語論文の検索や同性同名の区別に試行錯誤する学生が多く見られた。
- ・Web of Scienceの絞り込みでは、分野や所属での絞り込みが難しく、フルネームで検索し、答えを出している学生も多かった。

新聞 DB でもどうやって絞り込んだらよいのかという質問もあったため、絞り込みの方法についてヒントをまとめて配布してもよいかもしれない。

- 考察については 1 つ 1 つ項目立てたことはよかったが、その解答の精度については個人でばらつきがあったため、解答例について検討が必要である。
- Web of Science の使用中に、ECS-ID・パスワードを正しく入力しても認証されないことがあった(電子ジャーナル・データベース認証システムの移行期間だったためか?)。メディアセンターの端末の IE では[インターネットオプション]が表示されないため、[セーフティ]閲覧履歴の削除でキャッシュのクリアを行う必要がある。
- 学生の感想では、「各データベースによって利点欠点があり、ひとつのデータベースだけを頼るのではなく複数のデータベースを参照することでより情報を得やすくなると思う」や「今まではレポート作成の際にグーグルで検索した内容や適当な書籍を用いるのが常だったが、データベースを活用することで実用的な調査の幅が広がったように思う。」などが寄せられ、講義・演習の効果が伺われた。
- Panda での課題提出および採点については、特に問題がなかった。

(担当:インターネット・データベース演習担当チーム)