

第89回
(2016.6.7, 6.14)

『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(6月7日)

・配布物:スライド

http://www.kulib.kyotoru.ac.jp/modules/support/content/2016tansaku/INDB_20160607.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。
- ・学術情報とは?: 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。
- ・電子書籍(Kindle/Pad): 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。
- ・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。
- ・学術誌のステータス: Impact Factor 該当雑誌1論文あたり発表後2年間での被引用数の平均値。個々の論文の影響度ではなく、雑誌全体の平均値。Journal of Citation Reportにて調べられる。
- ・論文の影響度を計る指標の多様化: Altmetrics Twitterなども要素に加えた、論文の影響度の指標。
- ・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
- ・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(インデックス作成)により高速な検索が可能になった。

○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)、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。

・データベースとシソーラスから始まる研究: 米国の有害事象データベース FAERS のデータ解析による研究

○しくみから考えるネット検索での落とし穴

・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、永続性が低い。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

○演習に向けて

・研究業績検索のポイント

:必ず複数の情報源をあたって比較する

:本当にその人かどうか確かめる

:発信者は誰で、いつの情報か確かめる

:重要な情報は何か考えて整理する

:情報が少ない場合の理由を考えてみる

:ネット検索もしてみる

■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、情報が少な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

:情報が少な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(6月14日)・・・受講者数:22名

◇演習ページ URL: <http://www.kulib.kyotou.ac.jp/support/12339>

■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: データベースや検索語の適切な選択を体得し、情報が見つからない場合の考察を行う。

○注意点: 同姓同名の人物を見分ける際、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. 2006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共著者も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2006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 同姓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)

- 藤田 和生 (文学研究科)
- 鈴木 基史 (法学研究科)
- 文 世一 (経済学研究科)
- 石原 昭彦 (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)

◇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

- ・ 選択する教員を変更した。

◇感想・反省点

- ・ 課題提出者は21名、平均点12.47点(15点満点)、得点率83.1%。
- ・ CiNiiで、論文検索ではなく著者検索をして被引用数を質問した人が多かった。来年度はうまく論文検索に誘導できればよい。
- ・ 英語論文の検索や同姓同名の区別に試行錯誤する学生が多く見られた。
- ・ 石原先生を選択して回答した学生の多くが、3-2で別の論文を挙げていた。イニシャルではなく、フルネームで検索したからだと思われる。
- ・ 課題3について、先生本人の特定に成功し、論文の記述もあっているが、対象となる論文が間違っている場合は部分点を与えた。

- CiNii Articles や Web of Science は著者情報の統制がしっかりとられていないため、Scopus など、著者情報が統制されているデータベースを一つくらい使ったほうがよいかもしれない。
- 学生の感想では、「一人の研究者がどんな人であるかを調べるのにも、様々な角度からの検索が可能であることが分かった。」や「データベースごとの特性を理解し、将来的に使いこなせるようにしていきたい。」などが寄せられ、講義・演習の効果が伺われた。
- アンケートでは、授業の難易度について難しかったと回答する割合が高かったが、有用度平均値も高く、役に立つと感じている学生が多いことがわかった。
- 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)