

第89回
(2018.5.29,6.5)

『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(5月29日)...受講生数:18名

・予習ビデオ

<https://www.youtube.com/watch?v=um5mOxgiNyo>

◇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ことができる。

(開始前)

前回授業の際、YouTubeにあげた金子先生作成の動画を見て予習してくるよう案内。

当日はあらかじめこちらで決めておいたグループ(1グループ2~3人、全6グループ、学部がばらけるように)ごとに座ってもらう。

(開始後)

まず、予習ビデオを見てきたかどうか、PandAを使って小テスト。(5点満点:7人、4点:10人)

◇講義内容(45分) 予習ビデオに肉付けした内容 PandAにスライドアップ。配布はなし。

○学術情報とインターネット

- ・**研究における情報の重要性:** 研究の価値は「新しい」こと。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。
- ・**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。
- ・**情報の検索にあたってのスキル:** 一般の情報検索は、網羅的でなくても多くは問題ないが、先行研究の情報検索は、見落としがあるのはまずい。十分かつ最新の情報を探し出すために、まず、図書、学术论文、データベースなど情報源の特性を理解する。そして、検索ワードや絞り込みなど検索技術を磨く。
- ・**学術資料:** 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、二次資料として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。
- ・**資料の特性**
 - :書籍・学術雑誌 一次情報、査読された内容、電子化の進展、永続性がある
 - :データベース 二次情報、分類や索引づけされ検索しやすい、永続性がある
 - :インターネット(Webページ、動画・画像) 不確かで間違った情報も含む、最新情報を掴みやすい、永続性はない
- ・**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、検索できないという意味がない、引用等がリンクできると便利、等の特性から、学術情報は早く電子化されるようになった。
- ・**学術誌のステイタス:** 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。問題点もある。
- ・**論文の影響度を計る指標の多様化:** Altmetrics SNSなどでの言及も要素に加えた、論文の影響度の指標。

○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。
- ・**内容** 1億語のコーパス(英語論文抄録)を解析してつくった医学辞書。単語の前後の文も表示することで、単語の使い方が分かりやすい。

○しくみから考えるネット検索での落とし穴

- ・**言葉の「ゆれ」**
 - :事物を表す言葉にはゆれがある。日本語でも英語でも状況は同じ。
 - :検索語の単純一致では多くの情報を見逃す

:検索システムで語の「ゆれ」を吸収する仕組みが必要。Googleなど「賢い」システムでは実装されているが、すべての情報検索に実装されているとは限らない

・概念や事物を表すことば

- :情報検索は「ことば」で行われるから、用語法を知ることが大切
- :概念の範囲は場合によって曖昧なので、正しく限定する必要がある。(≒粒度)
- :同姓同名など、固有名詞の重複もよくある。
- :専門的なことは日本語だけの調査では不十分。

・シソーラス(同義語辞書)

- :概念をあらわす統制語を決め、同義語を寄せ集める。
- :概念の上位下位によりツリーをつくる。
- :概念同士の関係性を数値化し、一定形式で記述する。
- :人工知能に必要なリソース

・検索スキルを上げるために

- :複数の情報源を使いこなす
- :検索語を様々に変えて検索する
- :情報の信憑性や客観性を確かめる
- :複数キーワードや演算子で的確に絞り込む
- :重要な情報は何か考える
- :一次情報にまで遡って調べる
- :情報が得られないなら理由を考えて見逃しを防ぐ

○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、いつの情報が確かめる
- :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)
- :いつの情報が、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

◇グループワーク(30分)

「20世紀の課題」という大枠から、各グループで具体的なテーマを決めてもらい、京大または他大学の教員からその分野の専門家を選んでもらう。テーマと選んだ教員はグループごとに紙に書いて提出。参考として上記のデータベースをPandAで提示。

●第9回演習(6月5日)・・・受講者数:16名

◇演習ページURL:<http://www.kulib.kyotoru.ac.jp/support/12303>

■課題 グループで設定したテーマに沿って、京大に在籍する教員を1人選び、下記のデータベースを利用して研究業績を調査せよ。(京大に適切な教員がいなければ、他大学の教員でも構わない)

なお、レポートには下記の調査項目について記載すること。

○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. 2008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2008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 (Clarivate Analytics)
 - ・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)
 - ・日経テレコン(日本経済新聞・京都新聞)
- (新聞データベースについて、6/5-12は同時アクセス数を増やした。)

◇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使用の案内、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

(第8回)

- ・講義前にYouTubeにあげた動画を見て予習してもらい、授業冒頭で確認のための小テストを行った。
- ・分類チームでのグループ分けを意識しつつ、総合演習チームの意見も取り入れ、INDBの回でグループ分けを行った。理系の学部のみまとめ、それ以外は学部がバラけるように決め、1チーム基本3人とした。調査する教員については、グループで統一しなくてもよいことにした。

(第9回)

- ・使用するデータベースとして日経テレコンを追加した。

◇感想・反省点

- ・ 課題提出者は16名、平均点11点(15点満点)、得点率73.3%。
- ・ 確認の小テストの結果からも予習ビデオの効果が高いことが分かった。
- ・ 来年度は同姓同名の人物の絞り込みについてORCIDの話題を取り入れるか検討する必要がある。
- ・ テーマ設定について、もう少し具体的にテーマを絞れるよう工夫があった方がよいかもしれない。
- ・ 課題の文章について、わかりにくい点があったので改善の余地がある。

(担当:インターネット・データベース演習担当チーム)