

第 10-12 回
(2012.12.17、
2013.1.7、1.15)

『インターネット情報およびデータベースとその活用法』 金子周司教授(薬学研究科)

●第 10 回講義 (12 月 17 日)...受講生数:34 名

・配布物:スライド(抜粋) (スライド完全版と動画は講義終了後、機構 HP>情報探索入門のページにアップ)

<http://www.kulib.kyoto-u.ac.jp/modules/support/content/2012tansaku/20121217kaneko.pdf>(スライド)

<http://www.kulib.kyoto-u.ac.jp/modules/support/content/2012tansaku/20121217kaneko.mov> (動画)

◇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。

◇講義内容

○学術情報とインターネット

・**研究と文献**: 研究の価値は「世界初」であることが重要。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。

・**学術情報＝文献?**: 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。

・**文献検索**: 二次情報(抄録検索)から冊子体の一次情報へというのが基本だったが、学術雑誌の電子化(投稿・編集・出版・公開)により電子媒体の一次情報にアクセスすることが多くなった。

・**インターネットと学術情報**: 学術情報は本来インターネット向き(多種類、少部数、世界共通)。医学・生命科学では 90 年代に電子化(抄録 DB と電子編集・出版のリンクや遺伝子や物質 DB との統合検索)が開始。

・Entrez/PubMed 統合データベース

: 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館の書誌情報が融合

: 今では化合物や疾患も統合されている

: 統制語 MeSH が柔軟な検索を可能に

・情報検索における現実的問題

: 表記のゆれ: “たばこ” “タバコ” “煙草”

: 同義語: “後天性免疫不全症候群” “エイズ”

: 多義語(特に略語): “AV” “CD”

: 階層性: “京都” “京都府” “京都市”

: 英語と日本語: “vector” “ベクター” “ベクトル”

○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。

・学生にありがちなミス

: 学習辞書で学術論文を読もうとする

: 日本語で書いてから英訳する

: ネットで日本語の情報ばかり集める

: 同義語、表記ゆれ、粒度に無頓着

・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、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)。

・**応用例**: 電子カルテからの副作用情報の発見、オンデマンド英語教材、パソコン辞書(英和・和英辞書、マウスオーバー辞書)、iPod コンテンツ(耳で覚えるライフサイエンス英語)

・**LSD プロジェクト**: シソーラスから連想検索へ、テキストマイニングへの応用、Methodology は確立しているので他分野の専門辞書にも応用可能。

○演習のツボと落とし穴

■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: J-GLOBAL、Read & Researchmap、Web of Knowledge、KAKEN

・研究業績検索のポイント:

:複数の情報源をあたる

: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

:情報を整理する(重要な情報は何か) ※被引用回数やインパクトファクター(IF)などの指標も参考に
(ただしこれらの数字が大きければよいというものでもない)

:情報が無い場合の理由を考える

■インターネット演習: 課題の概要、目標および過去問題の紹介。情報やサイトの信頼性、検索語による結果の差異などを考察できるようになることが目標。

・英語の医学系論文を探す: 医学系の場合はPubMed(最新情報に向く、シソーラス・リンク充実)、Web of Science(正式な被引用状況を調べられる)、Google Scholar(手軽に代表的論文を調べられる)などデータベースの特長にあわせて使い分けるとよい。

・検索のポイント

:いつの時点の情報か

:正しい名前(用語)は何か

:どうキーワードを掛け合わせるか

:どこが情報の発信源か

:表記のゆれ、同義語、言い換え表現

:ネットには諸説入り乱れている

:科学的な情報=総説や論文を探す

:科学的情報は英語で検索する

:科学的情報に「定説」はない

●第11回演習(1月7日)・・・受講者数:31名

◇演習ページ URL: http://www.kulib.kyoto-u.ac.jp/modules/support/index.php?content_id=20

■例題1

雑誌「CELL」の2011年のインパクトファクターを求めよ。計算式も書くこと。

「CELL」に載った論文数:2009年=373本、2010年=319本

CELL掲載論文が2011年に他の論文に引用された回数:2009年=13,219回、2010年=9,204回(※被引用回数は2013/17時点のもの)

(+)÷(+)÷() (小数点第4位四捨五入)

○目的:実際に計算することでインパクトファクターへの理解を深める。

■例題2

2006年8月25日の『Cell』誌に掲載された山中伸弥氏の以下の論文の被引用数を調べなさい。

"Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors"

その際、下記のデータベース等の中から二つ以上を選んで被引用数を調べること。

また、それぞれの被引用数に違いがあれば、その理由についても考察しなさい。

・Web of Science

・PubMed

・SciVerse Scopus

・Google Scholar

○目的:例題1の復習をかね、検索結果の違いについて考察する。

■課題1 京大に在籍する1名の教員(金子周司教授は除く)について研究業績を調査せよ。

○目的:データベースファイル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。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 科学研究費助成金の獲得状況
3. 学位および学位論文
4. 最近の著作物、論文
 1. 最も被引用回数の多い英語の論文(被引用数が0の場合も含む。)
 2. 上記(4-1)で調べた英語論文が掲載されている学術誌の2011年のインパクトファクター
 3. (上記(4-1)で英語論文がヒットしない場合)最も被引用回数の多い日本語の論文か、被引用数がない場合は最近の論文
 4. (上記(4-1)で英語論文がヒットしない場合) 英語論文のない場合の考察

■使用するデータベース

- ・JGLOBAL
- ・ReaD & Researchmap
- ・NDL-OPAC (国立国会図書館)
- ・CiNii Articles・日本の論文をさがす(国立情報学研究所)
- ・Web of Science
- ・Journal Citation Reports
- ・SciVerse Scopus
- ・KAKEN 科学研究費助成事業データベース(国立情報学研究所)
- ・その他の情報源 (京都大学ホームページ、京都大学博士学位論文データベース、京都大学教育研究活動データベース、Google Scholar、先生個人や所属機関が開設しているホームページ等)

■課題2

以下に3つのDBの特徴、検索する際に注意すべきことを述べよ(各200字以内)。

ただしWeb of ScienceとCiNii Articlesは解答必須とし、下記の挙げられたDBの中から1つ選択して答えること。

- ・Westlaw Japan
- ・聞蔵IIビジュアルfor Libraries
- ・JDreamII
- ・医中誌 Web

★ヒント★

収録対象、収録年代、データ更新時期、検索記号、京大における同時アクセス数 etc...

○目的:それぞれのデータベースの概要を把握し、特徴を理解する。

■課題2解説資料 http://www.kulib.kyoto-u.ac.jp/modules/support/content/2012tansaku/DB_kadai2.pdf

●第12回演習(1月15日)・・・受講者数:25名

◇演習ページ URL: http://www.kulib.kyoto-u.ac.jp/modules/support/index.php?content_id=21

○例題

■例題1

1884年にボストンの Robert Brothers から出版された『宝島』(ロバート・ルイス・スチーブンソン著)の本文中に、登場人物の一人である「ジョン・シルバー」の名前が最初に登場するのは何ページか。

★ ヒント:日本語の題名や人名で検索しても見つかりません。英語で探してみましょう。

○目的:英語情報の検索をする。本の内容を見ることのできるサービスなどを活用する。

■例題2

日本の男性・女性それぞれの平均睡眠時間は何時間何分か?最新のデータを調べよ。

○目的:複数の情報を比較し、その中から信頼性の高い統計情報を選択する。

■課題

京都大学の派遣留学制度を利用して、ニュージーランドの大学に留学する場合の手続きを調査する。

1. 学内選考へ応募するために必要な書類を答えなさい。
2. 「1」の情報が掲載されているウェブページを答えなさい。
※ウェブページの題名、ウェブサイトの名称、URLの例
 - ・(ウェブページの題名)サービス案内
 - ・(ウェブサイトの名称)京都大学図書館機構
 - ・(URL)http://www.kulib.kyoto-u.ac.jp/modules/service/index.php?content_id=5
3. 大学間学生交流協定を締結している大学の中から、留学を希望する大学を1校選びなさい。
なお、英語の正式名称で答えること。
4. 「3」の大学の英語版公式ウェブサイト参照し、派遣留学(※一般留学は除く)の出願方法と必要書類を簡潔にまとめなさい。なお、必要書類は英語のまま引用してかまわない。
5. 「4」の情報が掲載されているウェブページを答えなさい。
6. 「3」の大学の英語版公式ウェブサイトと日本での公式窓口ページを比較し、それぞれのメリット・デメリットを述べなさい。
7. 「3」の大学の公式パンフレット(紙媒体の資料)が発行されていると仮定し、公式ウェブサイトとどのように使い分けると効率よく情報収集ができるか述べなさい。
8. 「1」-「7」において調べたサイト以外で、留学するにあたって役に立つと思うウェブサイトを1つ挙げ、そのサイトを評価する理由、提供されている情報の信頼性について述べなさい。
9. この課題に取り組んでの感想・反省を述べなさい。

○目的:探すターゲットの原語で、翻訳機能などを賢く使い検索する。紙とウェブ、それぞれの特性を知り、情報を補完したうえで信頼性を見極め、価値ある情報かどうか評価する。

■回答例 http://www.kulib.kyoto-u.ac.jp/modules/support/content/2012tansaku/m_answer2012.pdf

■ポータルサイトへのリンク

Yahoo! JAPAN(ヤフー)

goo(NIT レジナント)

Google(グーグル)

Excite(エキサイト)

Bing(マイクロソフト)

MARS FLAG(マーズフラッグ)

はてな(はてな)

担当:インターネット・データベース演習担当チーム)