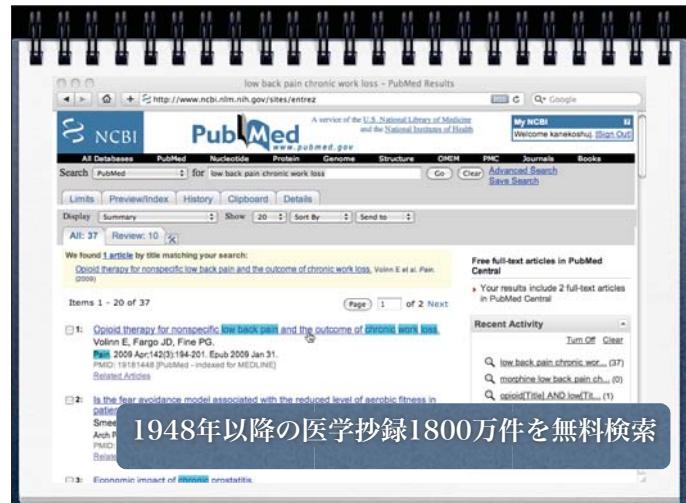


インターネットと学術情報

- 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
 - 遺伝子や物質DBとの統合



出所 : <http://www.ncbi.nlm.nih.gov/pubmed/>

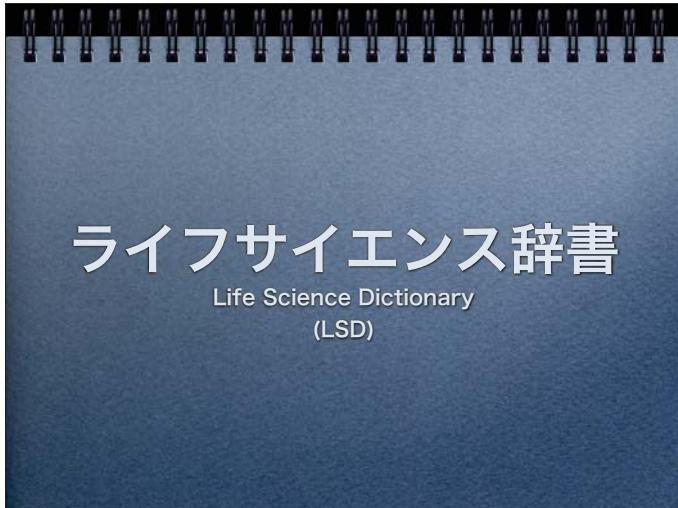


Entrez/PubMed 統合データベース

- アメリカでは医学図書館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
 - 結制誘 MoSH が柔軟な検索を可能に

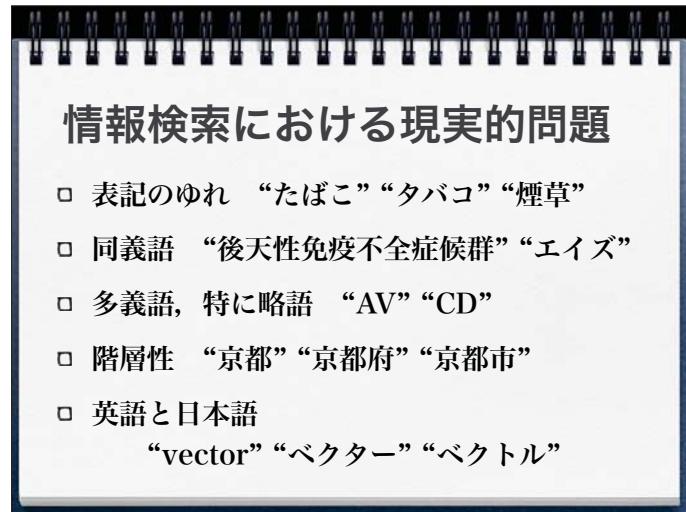


出所 : <http://www.ncbi.nlm.nih.gov/pubmed/>



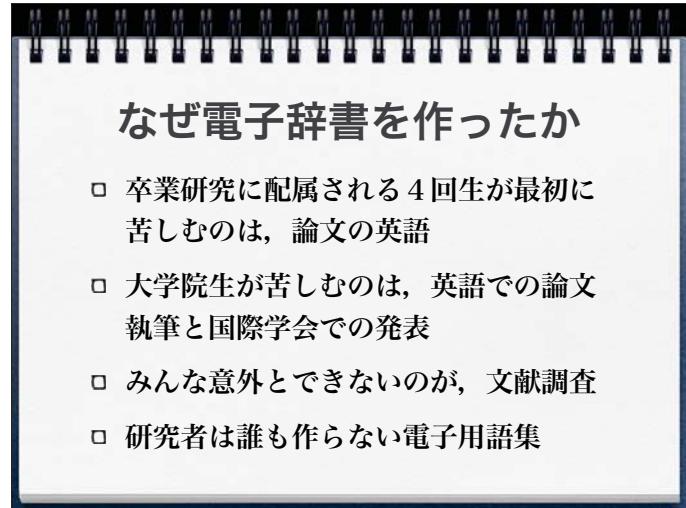
ライフサイエンス辞書

Life Science Dictionary
(LSD)



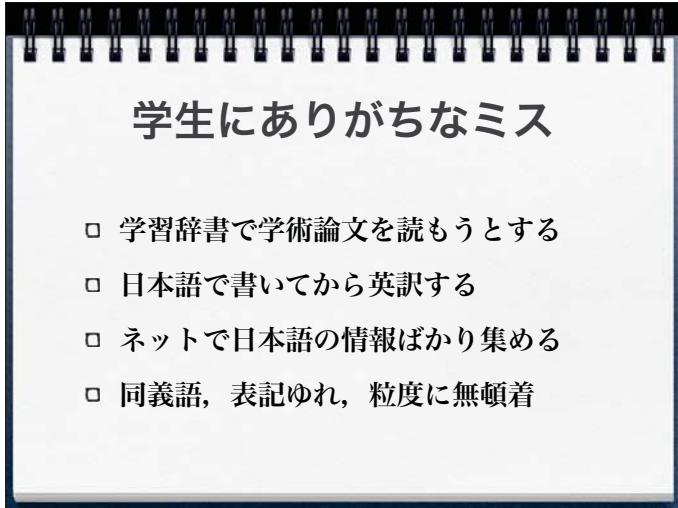
情報検索における現実的問題

- 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”

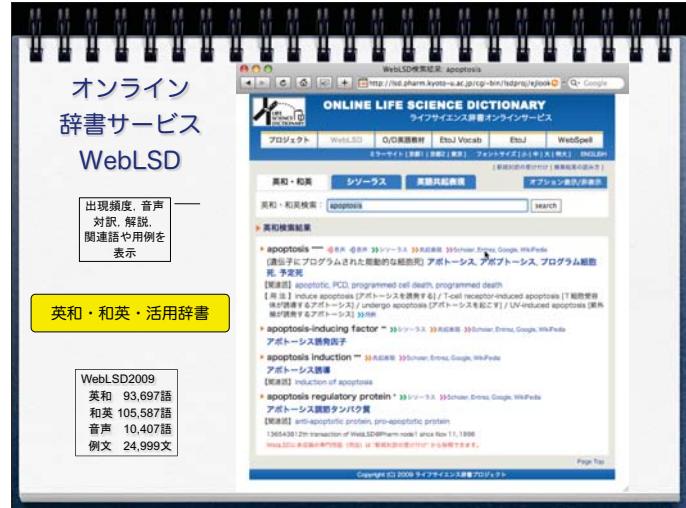


なぜ電子辞書を作ったか

- 卒業研究に配属される4回生が最初に苦しむのは, 論文の英語
- 大学院生が苦しむのは, 英語での論文執筆と国際学会での発表
- みんな意外とできないのが, 文献調査
- 研究者は誰も作らない電子用語集



- 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



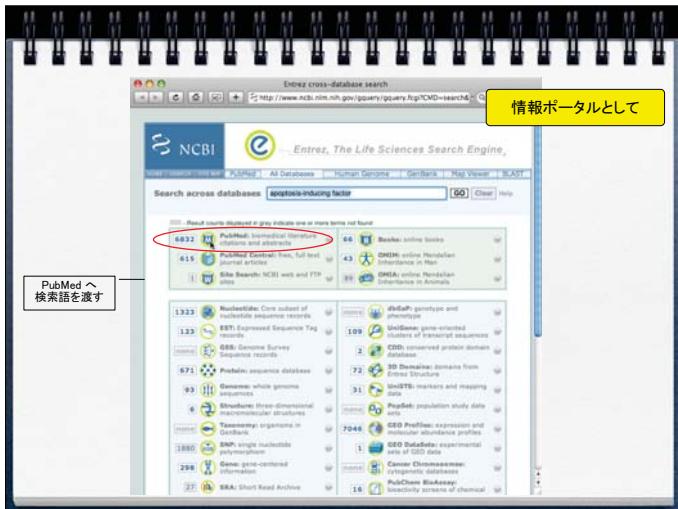
オンライン 辞書サービス WebLSD

出現頻度, 音声
対訳, 解説,
関連語や用例を
表示

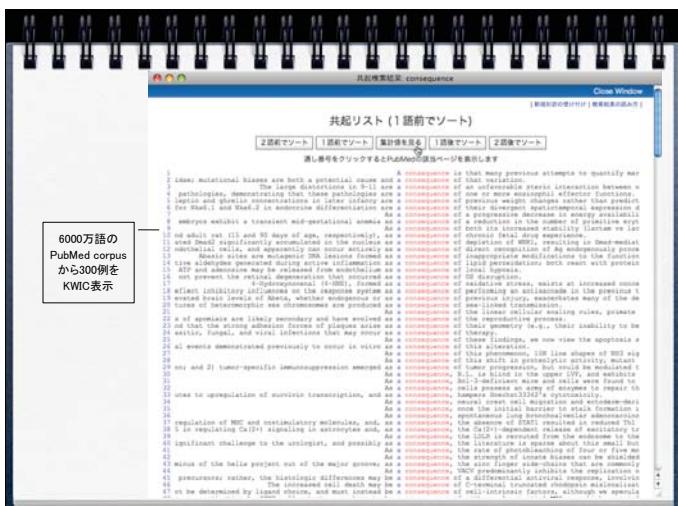
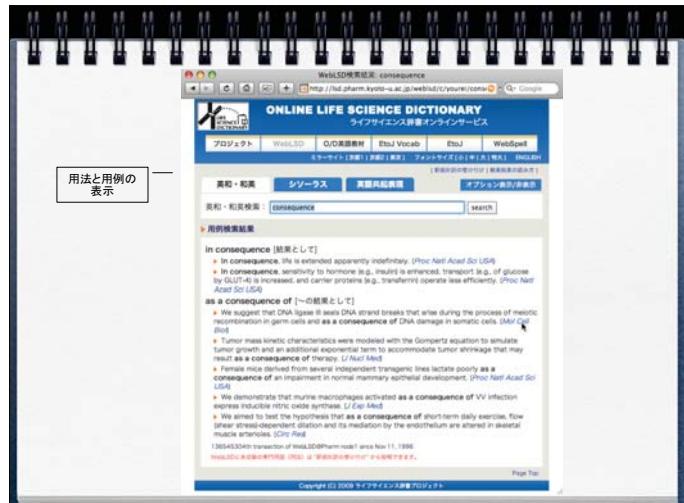
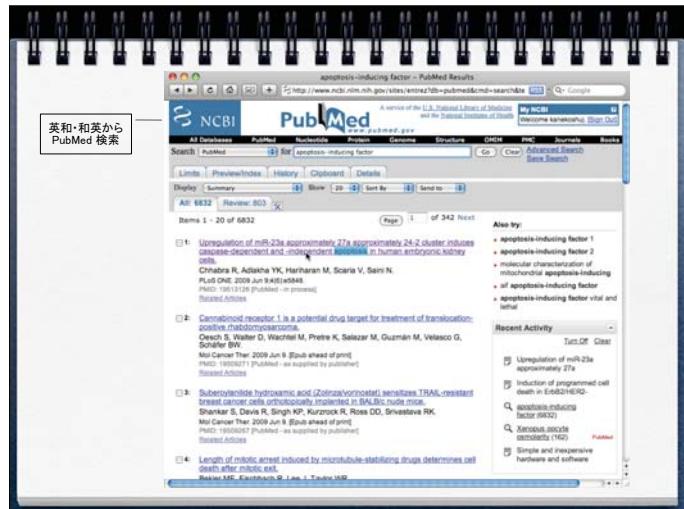
英和・和英・活用辞書

WebLSD2009
英和 93,697語
和英 105,587語
音声 10,407語
例文 24,999文

Copyright © 2009 ライフサイエンス辞書プロジェクト



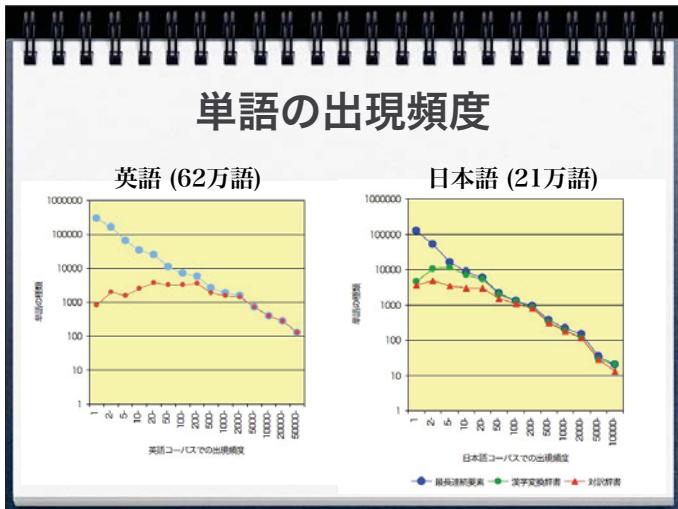
出所：<http://www.ncbi.nlm.nih.gov/>



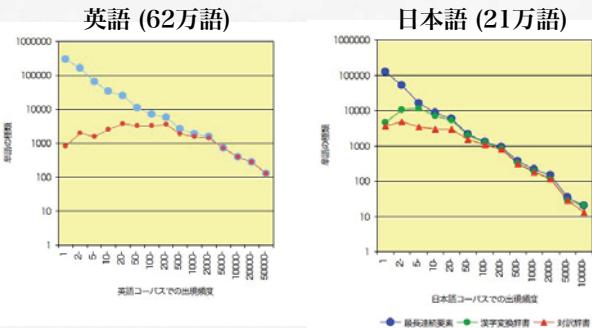
6000万語の
PubMed corpus
から300例を
KWIC表示

どうやって作ったか

1. 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



単語の出現頻度



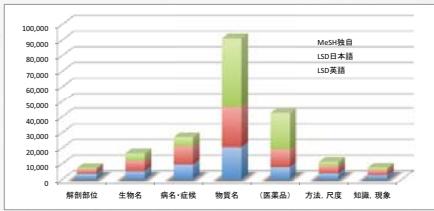
専門用語の冗長性

- 事物の名称は数多くの同義語をもち、表記のゆれも大きい
 - これらを英和1対の統制語とその他の同義語（シノニム）に整理する
 - 統制語には頻度の高い表記をできるだけ採用する
 - 遺伝子：カプサイシン受容体、バニロイド受容体1型、TRPV1 チャネル、capsaicin receptor, vanilloid receptor type 1, VR1 protein, transient receptor potential vanilloid 1, TRPV1 channel
 - 医薬品：プログラマ、プロトピック、Prograf, Propotic、タクロリムス（水和物）、 tacrolimus (hydrate), FK(-)506, FR(-)900506
 - 疾患：1型糖尿病、1型糖尿病、インスリン-[インシュリン]依存性[型]糖尿病、自己免疫性糖尿病、diabetes mellitus type 1, type 1 DM, insulin-dependent diabetes mellitus, NIDDM



LSDシソーラス (概念ツリー+同義語辞書)

- 17万語の専門用語を2.5万対訳の統制語に集約
 - 解剖部位, 生物名, 病名, 生体分子, 医薬品, 方法, 現象を網羅
 - うち2万語はMeSHツリーで上位下位関係を整理
 - MeSH 未収録の専門用語も多い (統制語で1万以上)



関連性の算出

- 文献抄録の中で最長一致する専門用語にタグ付け
 - タグ(統制語) 同士の共起頻度をカウント



2017-2018

ONLINE LIFE SCIENCE DICTIONARY

- ▶ **和英接続結果**

■ 糖尿 \rightarrow とうりゅう \rightarrow [Schoolar](#), [Entries](#), [Google](#), [WikiPedia](#)
 \triangleright [glycousuria](#) \rightarrow とうりゅう症
 \triangleright [glucosuria](#) \rightarrow とうりゅう症

■ 糖尿病 \rightarrow とうりゅうびょう \rightarrow [Schoolar](#), [Entries](#), [Google](#), [WikiPedia](#)
 \triangleright [diabetes mellitus](#) \rightarrow とうりゅう症 \rightarrow とうりゅう症
 \triangleright [diabetes](#) \rightarrow とうりゅう症
 \triangleright [diabetes DM](#) \rightarrow とうりゅう症
 \triangleright [diabetic](#) \rightarrow とうりゅう症 \rightarrow とうりゅう症
 \rightarrow とうりゅう症性糖尿病, とうりゅう症性高血糖

■ 糖尿病患者 \rightarrow とうりゅうびょうしゃ \rightarrow [Schoolar](#), [Entries](#), [Google](#), [WikiPedia](#)
 \triangleright [diabetic patient](#) \rightarrow とうりゅう症患者
 \triangleright [diabetic patient](#) \rightarrow とうりゅう症患者

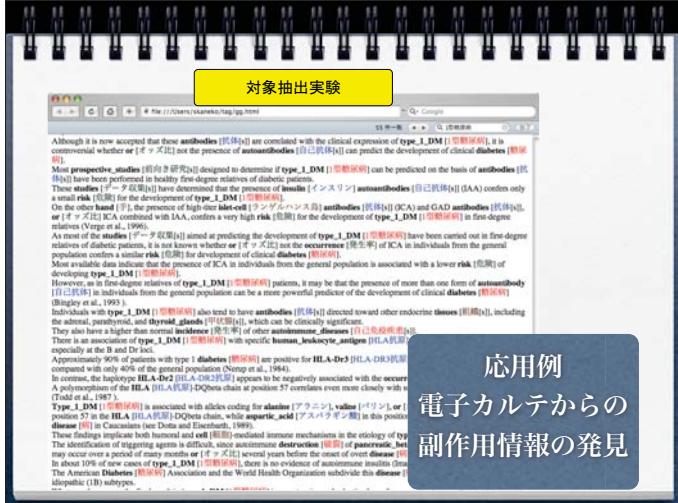




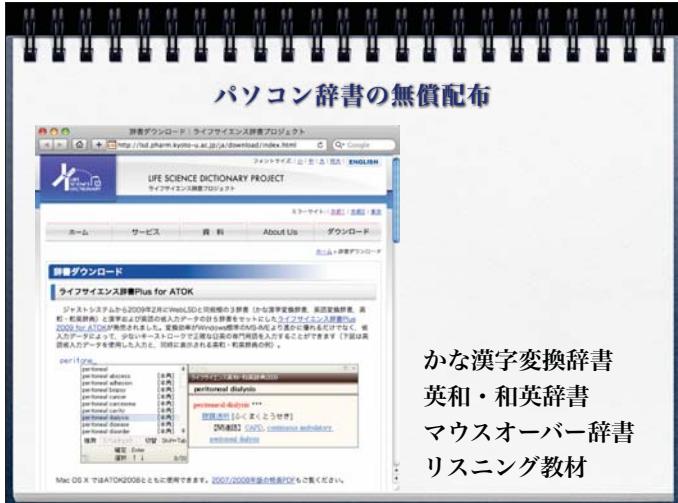
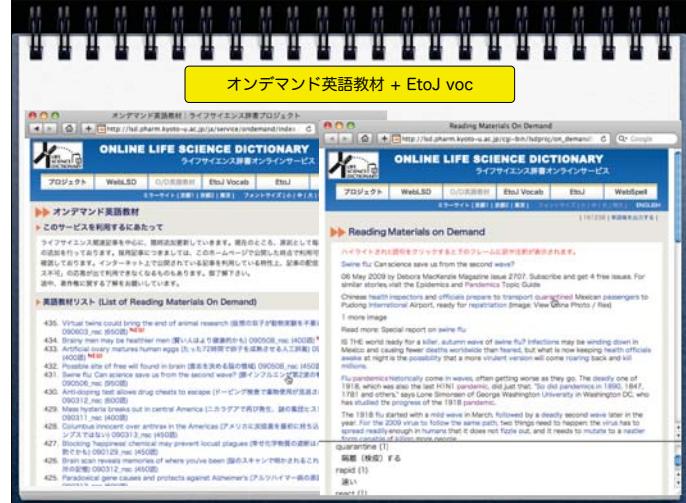
出所 : <http://www.ncbi.nlm.nih.gov/>



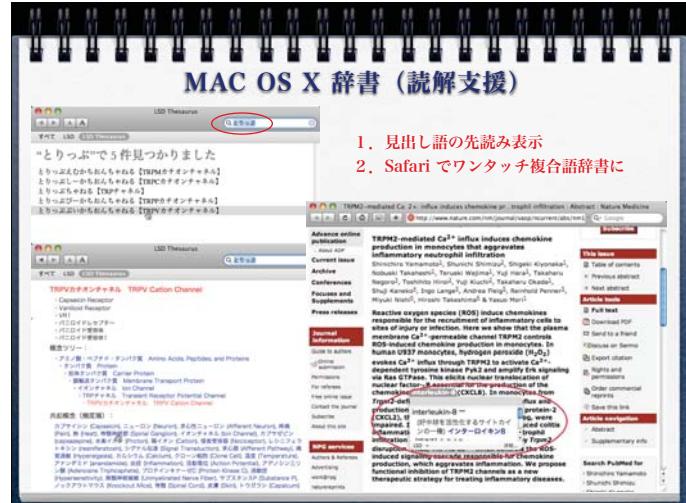
出所 : <http://www.google.co.jp/>

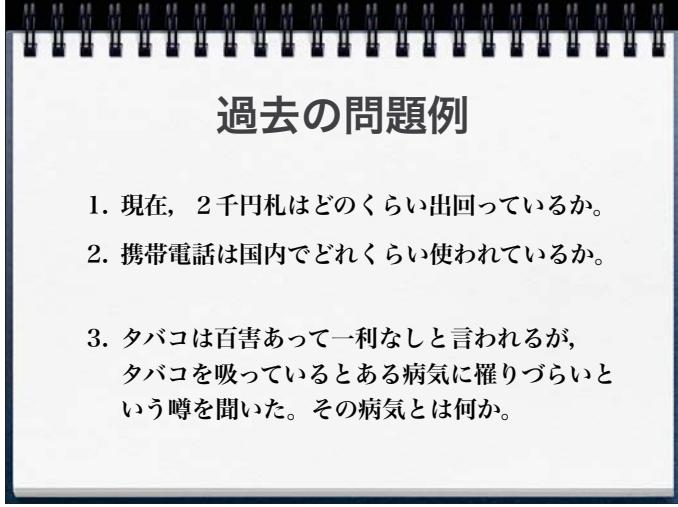
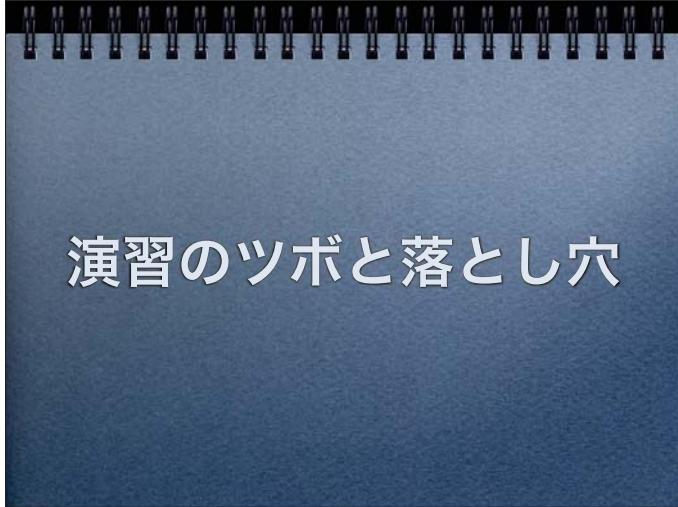


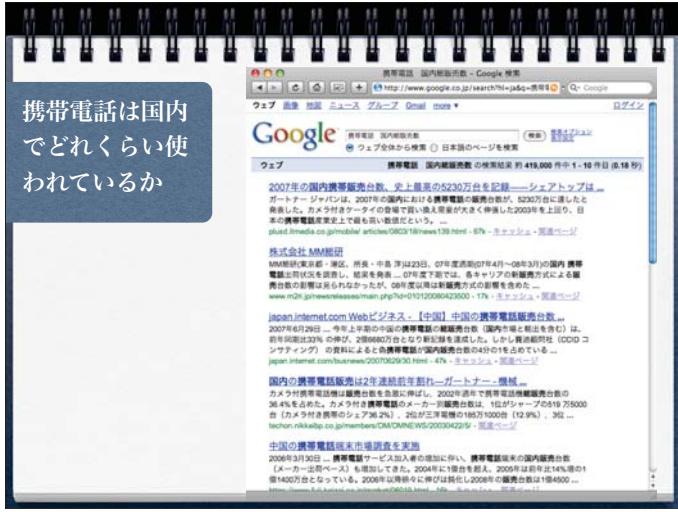
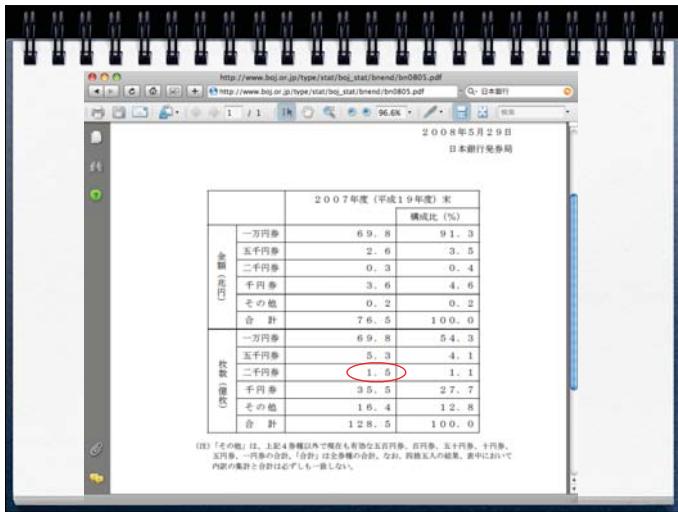
応用例
電子カルテからの
副作用情報の発見

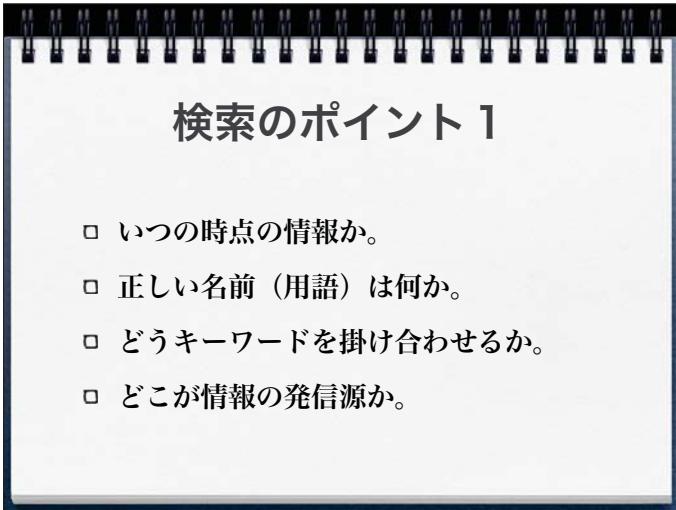


かな漢字変換辞書
英和・和英辞書
マウスオーバー辞書
リスニング教材

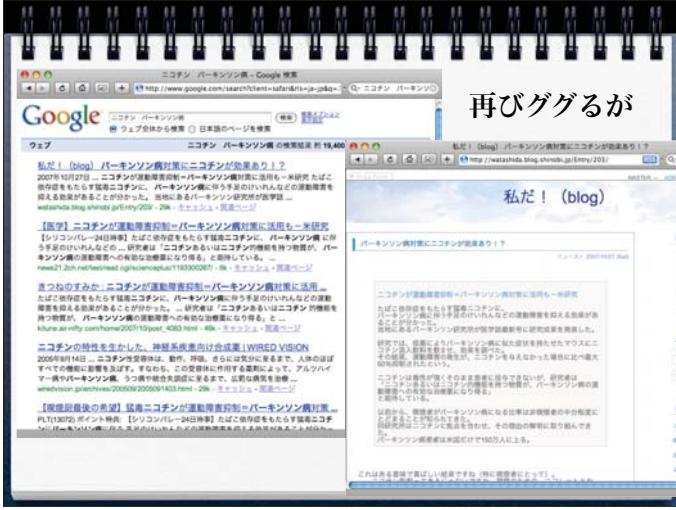
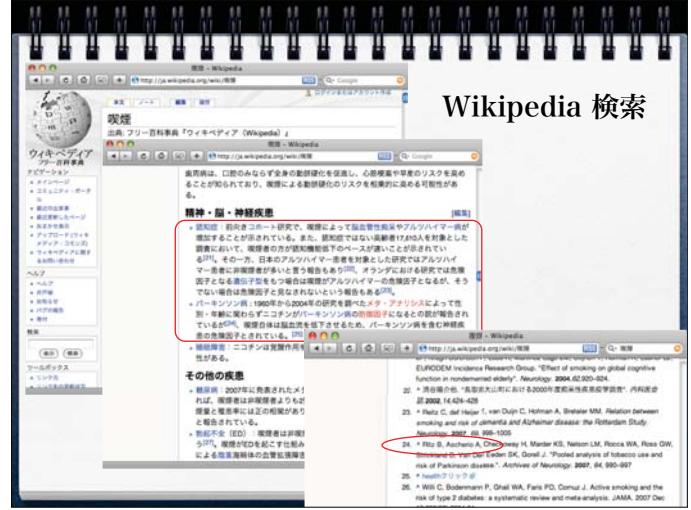
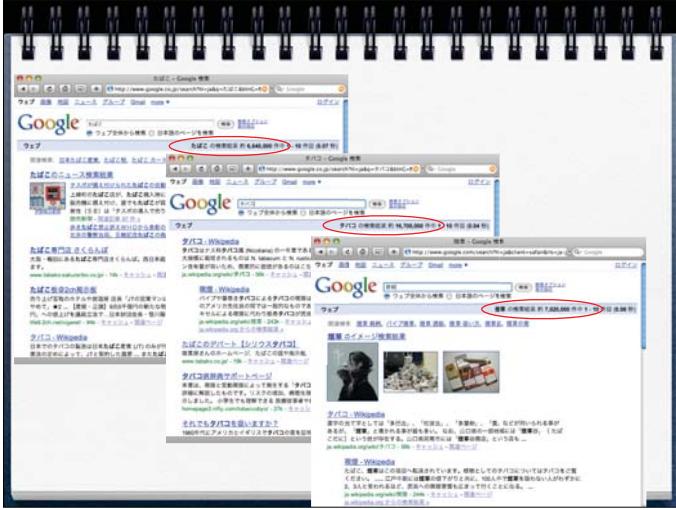
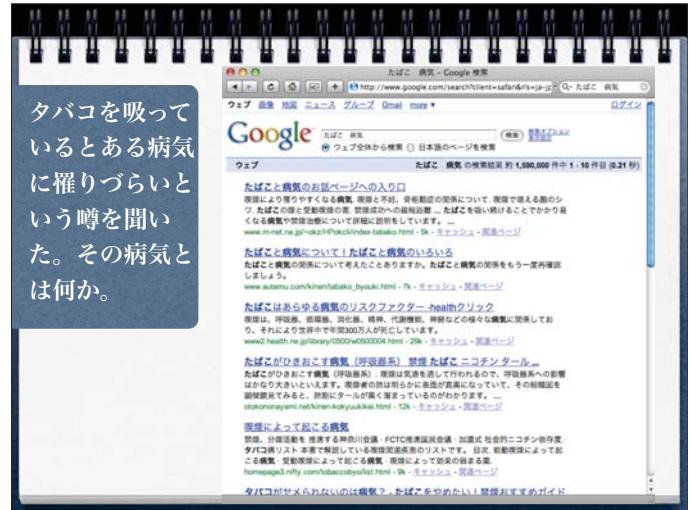


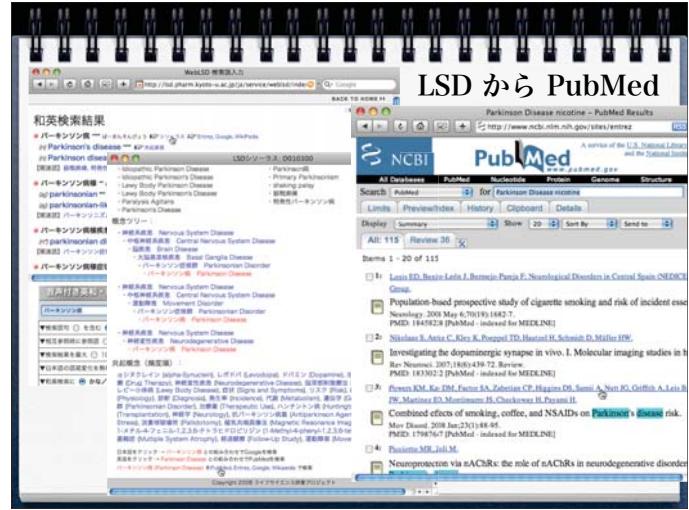
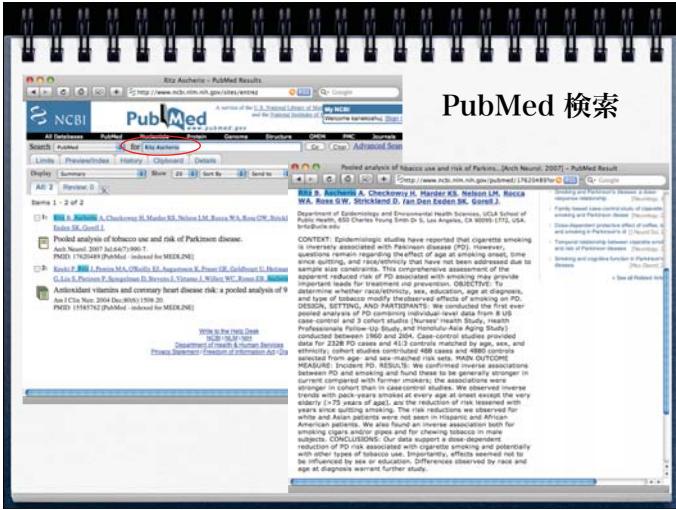
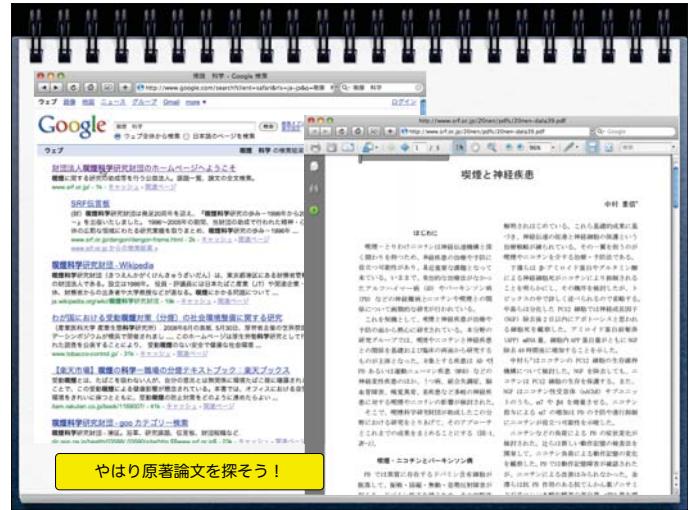
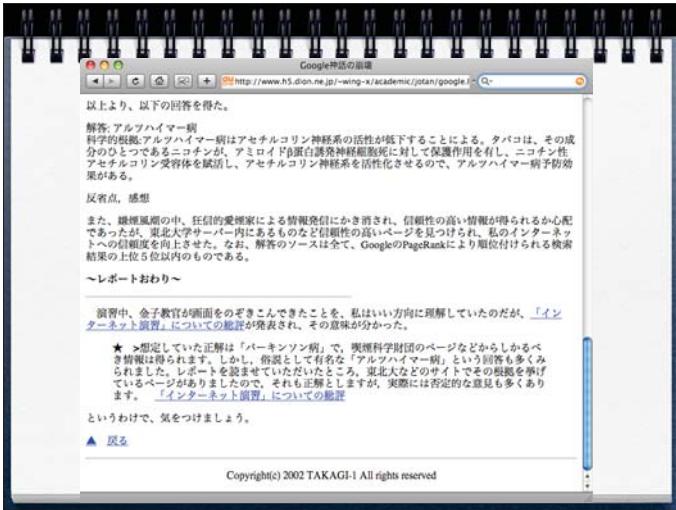
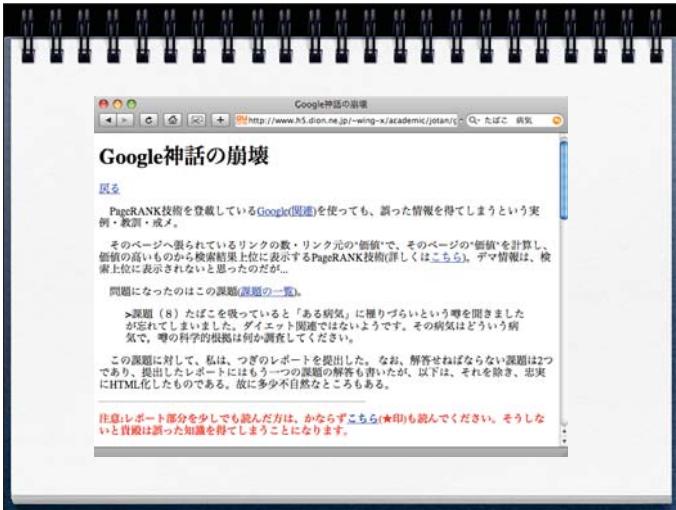


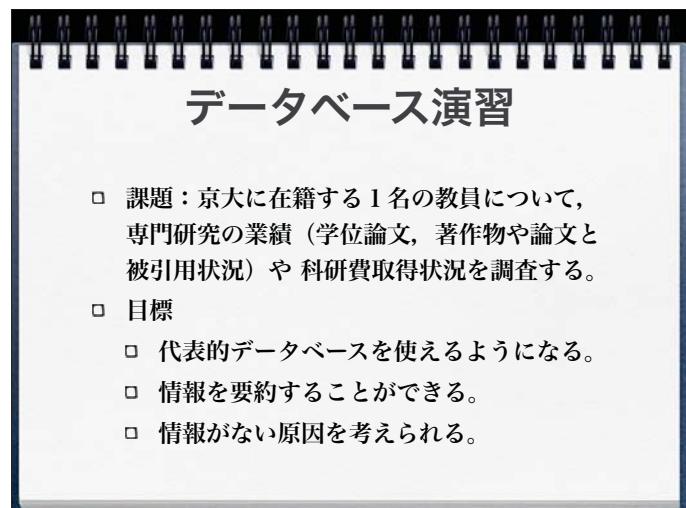
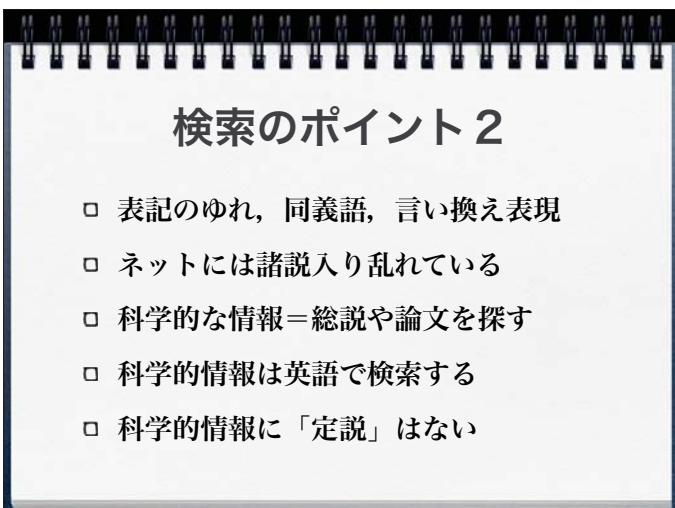
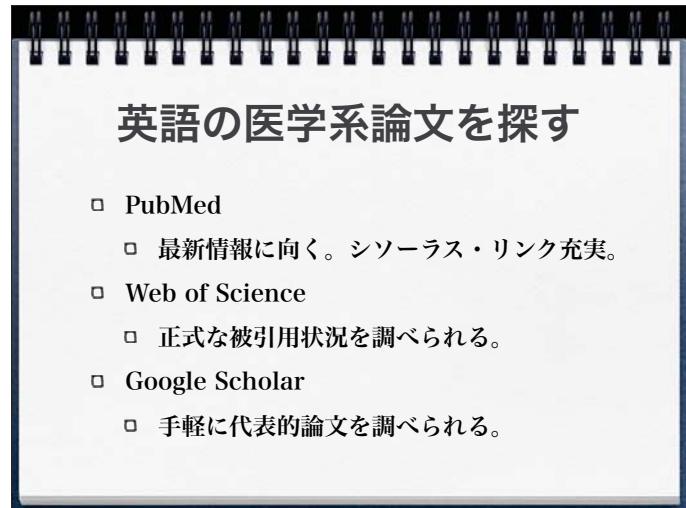
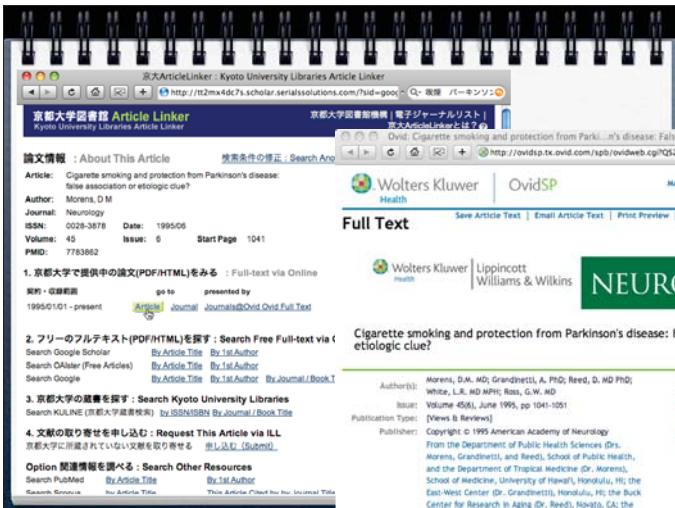




- いつの時点の情報か。
- 正しい名前（用語）は何か。
- どうキーワードを掛け合わせるか。
- どこが情報の発信源か。







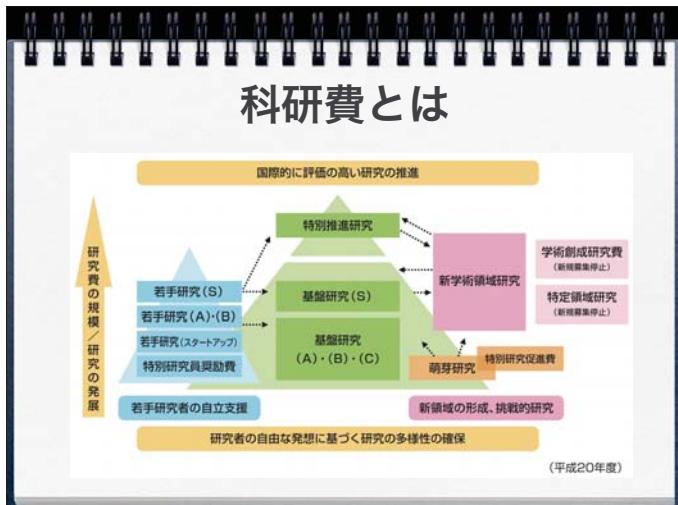
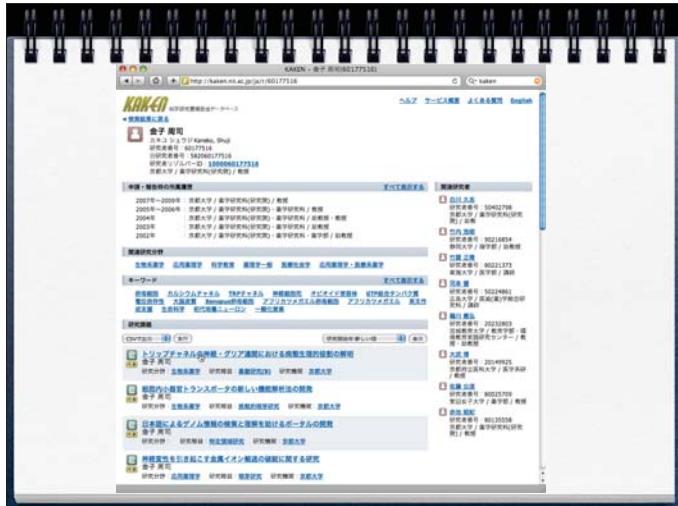
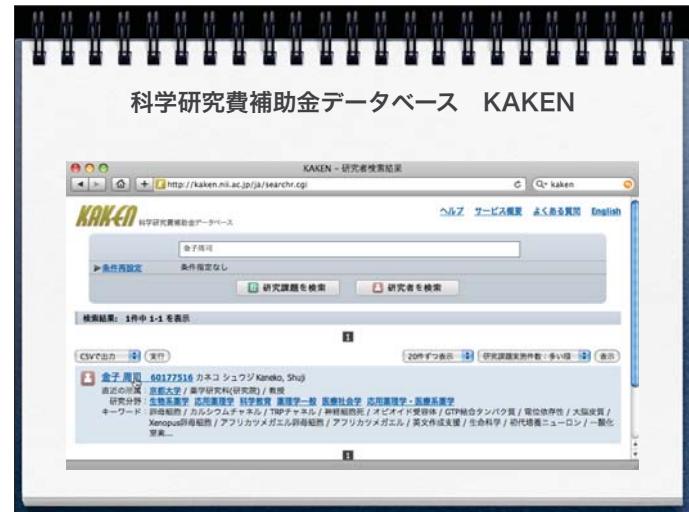
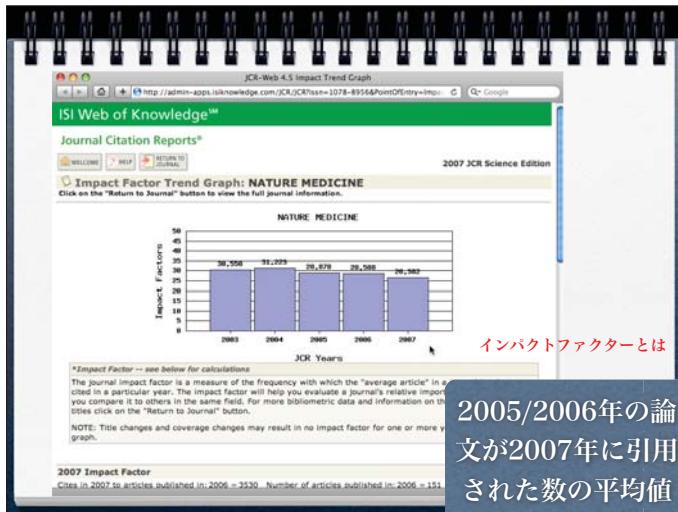
ReaD 研究者ディレクトリ (JST)

出所: <http://read.jst.go.jp/>

Web of Knowledge 被引用文献データベース

被引用回数が多い
論文を選ぶ

所属を確認
IF調査



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
- 情報を整理する（重要な情報は何か）
- 情報がない場合の理由を考える

おわり