

# インターネット情報とデータベースの活用法

情報探索入門

薬学研究科 金子周司

## 本日のメニュー

- 学術情報とインターネット
- ライフサイエンス辞書の紹介
- 演習のツボと落とし穴
- スライドは授業ページでPDFを公開

あんた誰？



## 研究と文献

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

研究の価値は  
「世界初」

## 学術情報＝文献？

- 書籍
- 学術雑誌
- データベース
- ウェブページ

永続性は保証される

永続性は保証されない

## 文献検索

一次情報  
(冊子体)

二次情報 (抄録検索)

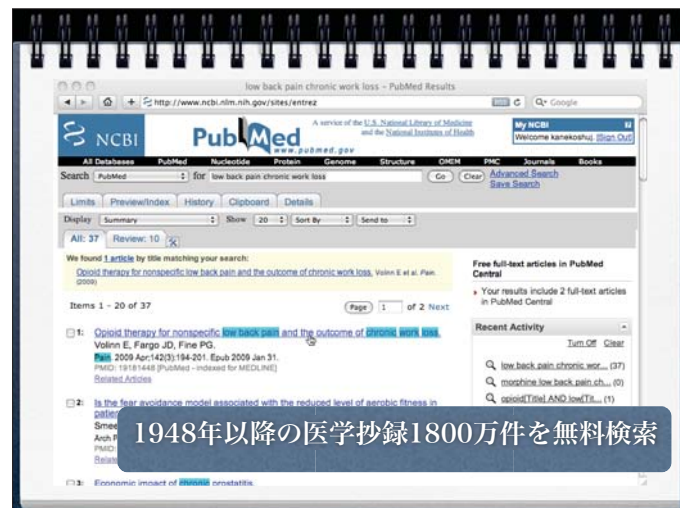


一次情報  
(電子出版)



# 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, 少数数, 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
- 統制語 MeSH が柔軟な検索を可能に





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

## 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

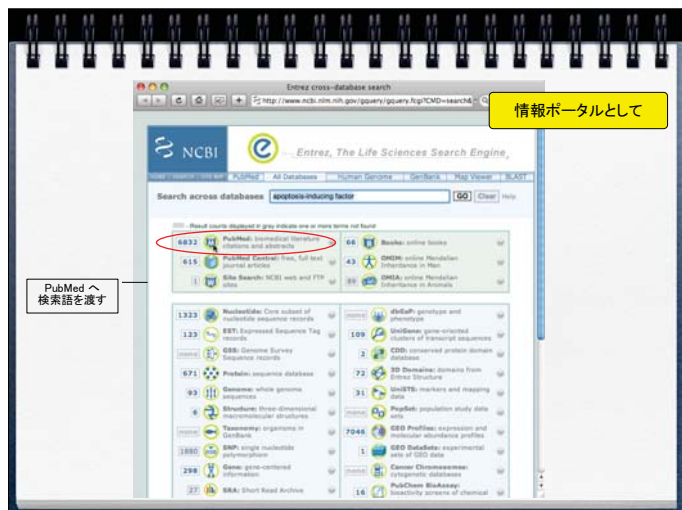
## オンライン辞書サービス WebLSD

出現頻度, 音声  
対訳, 解説,  
関連語や用例を  
表示

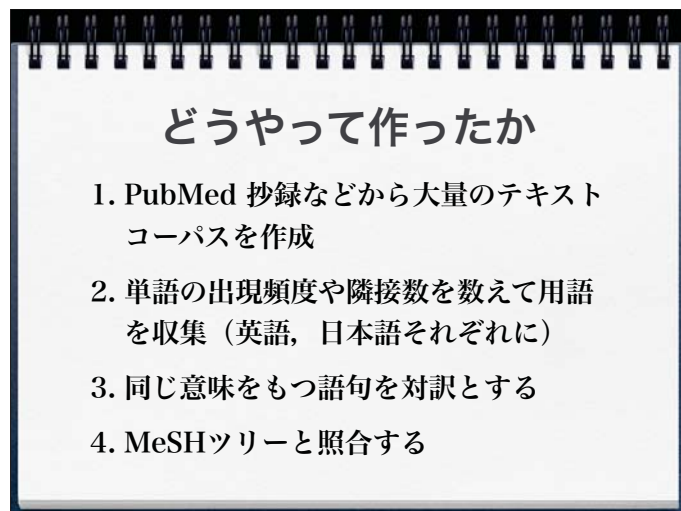
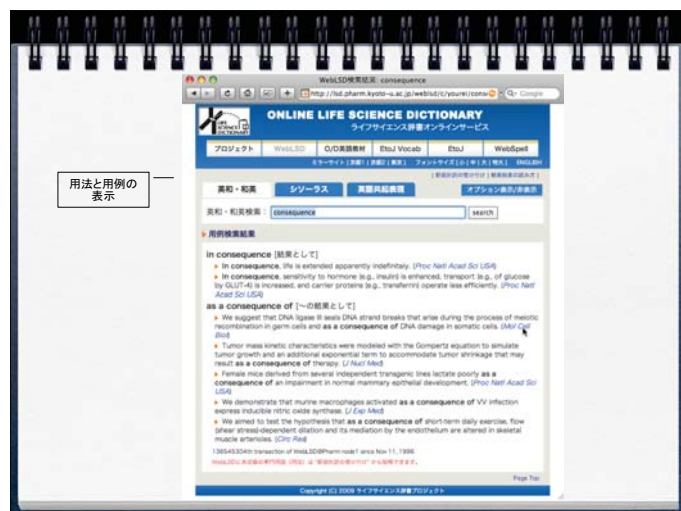
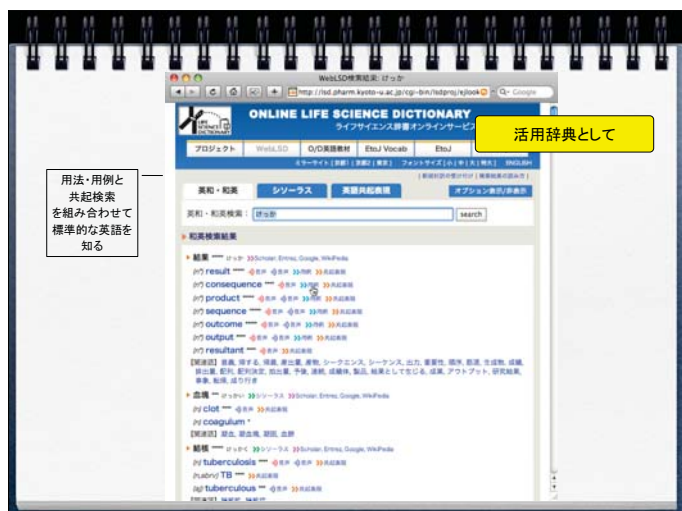
英和・和英・活用辞書

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



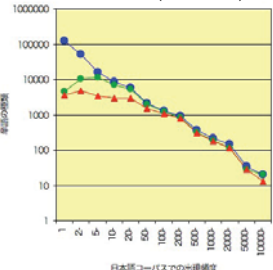
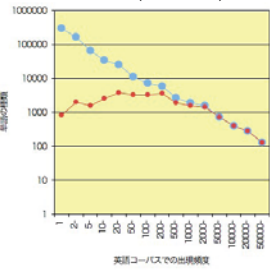
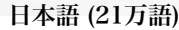


出所: <http://www.ncbi.nlm.nih.gov/>





## 単語の出現頻度

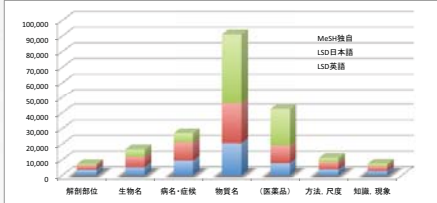


## 専門用語の冗長性

- **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 チャンネル，capsaicin receptor, vanilloid receptor type 1, VR1 protein, transient receptor potential vanilloid 1, TRPV1 channel
- **医薬品**：プログラフ，プロトピック，Prograf, Protopic, タクロリムス（水和物），tacrolimus (hydrate), FK(-)506, FR(-)900506
- **疾患**：I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

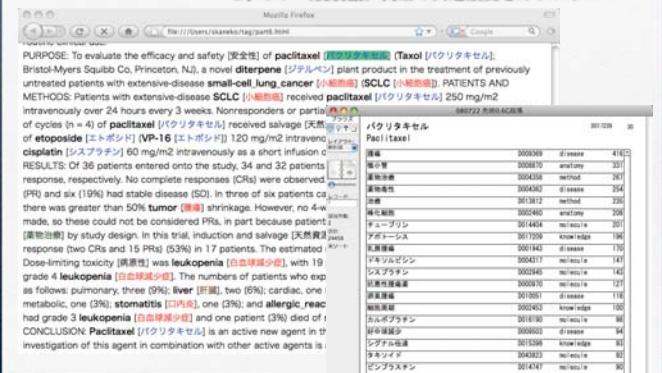
**LSDシソーラス**  
(概念ツリー+同義語辞書)

- 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

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



シソーラスとして

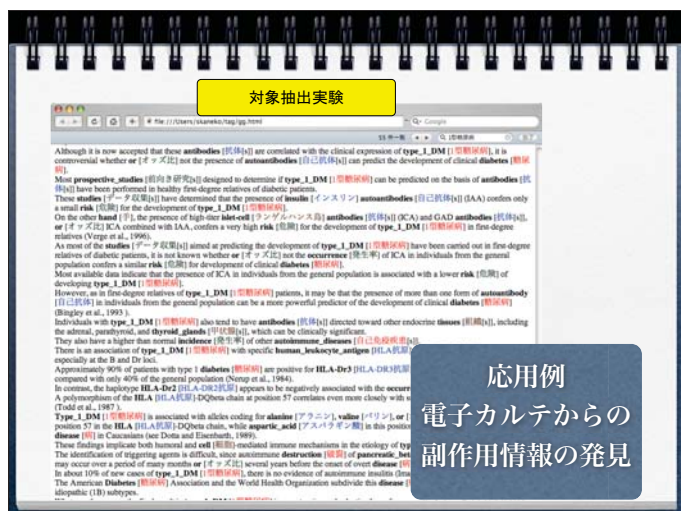




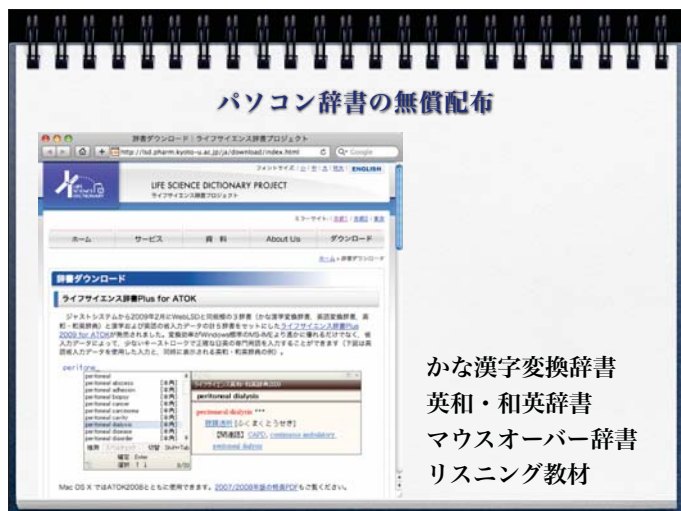
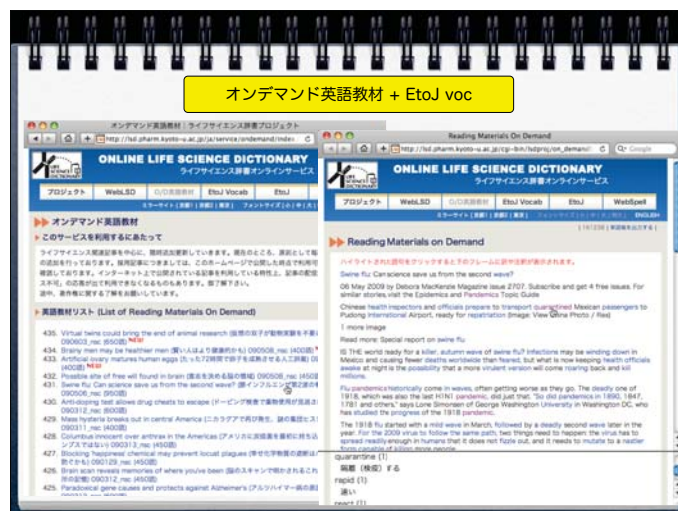
出所: <http://www.ncbi.nlm.nih.gov/>



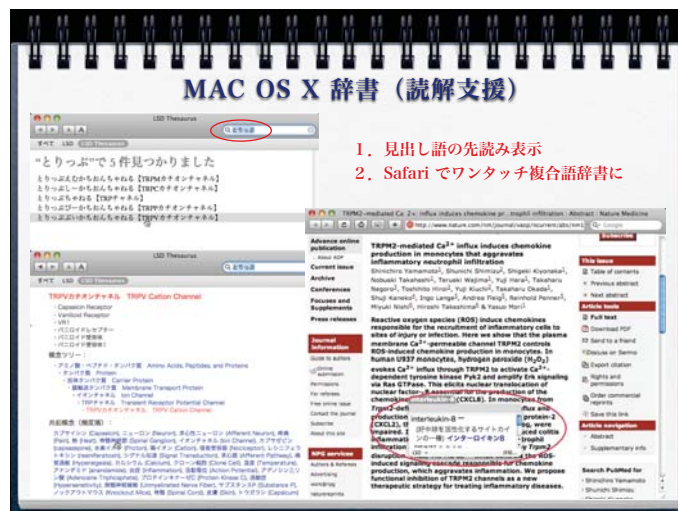
出所: <http://www.google.co.jp/>



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



1. 見出し語の先読み表示  
2. Safari でワンタッチ複合語辞書に



# 耳で覚えるライフサイエンス英語

必須2000語を iPod で



1. シャッフルしてヒアリング

2. 表示で確認

# LSDプロジェクト

- コーパスに基づく専門用語の分析
- 電子辞書を情報ポータルに
- 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

自分の専門分野で作りたい人、協力します

# 演習のツボと落とし穴

# インターネット演習

- 課題：演習日に授業ページで提示する7つの課題から1つ以上を自由に選び、インターネットを用いて情報を収集する。
- 目標
  - 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。

# 過去の問題例

1. 現在、2千円札はどのくらい出回っているか。
2. 携帯電話は国内でどれくらい使われているか。
3. タバコは百害あって一利なしと言われるが、タバコを吸っているとある病気に罹りづらいという噂を聞いた。その病気とは何か。





出所：http://www.boj.or.jp/



2008年5月29日  
日本銀行発券局

2007年度（平成19年度）末			
	金額（兆円）	構成比（%）	
一万円券	69.8	91.3	（注）「その他」は、上記4券種以外で残高も有効な五円券、百円券、五十円券、十円券、五円券、一円券の合計。「合計」は全券種の合計。なお、残高は入札の結果、表中において内訳の集計とは必ずしも一致しない。
五千円券	2.6	3.5	
二千円券	0.3	0.4	
千円券	3.6	4.6	
その他	0.2	0.2	
合計	76.5	100.0	
一万円券	69.8	91.3	（注）「その他」は、上記4券種以外で残高も有効な五円券、百円券、五十円券、十円券、五円券、一円券の合計。「合計」は全券種の合計。なお、残高は入札の結果、表中において内訳の集計とは必ずしも一致しない。
五千円券	5.3	4.1	
二千円券	1.5	1.1	
千円券	35.5	27.7	
その他	16.4	12.8	
合計	128.5	100.0	

年末銀行券発行高（2005年末）

金額（兆円）	平成16年（2004年）末		平成17年（2005年）末		前年比（%）
	金額（兆円）	構成比（%）	金額（兆円）	構成比（%）	
一万円券	70.2	90.1	72.1	90.9	+1.7
五千円券	2.8	3.6	2.8	3.6	
二千円券	1.0	1.3	0.5	0.6	
千円券	3.8	4.9	3.7	4.7	
その他	0.2	0.2	0.2	0.2	
合計	78.0	100.0	79.3	100.0	
一万円券	70.2	51.8	72.1	53.8	-1.1
五千円券	5.5	4.1	5.6	4.2	
二千円券	5.0	3.7	2.5	1.9	
千円券	38.1	28.1	37.2	27.8	
その他	16.6	12.3	16.5	12.3	
合計	135.5	100.0	134.0	100.0	

携帯電話は国内でどれくらい使われているか

2007年の国内携帯電話台数、史上最大の5230万台を記録——シェアトップは、ガートナー・ジャパンは、2007年の国内における携帯電話の総台数が、5230万台に達したと発表した。カマクラ・ジャパンの発表で、国内最大の携帯電話台数を記録した2007年を上回り、日本の携帯電話台数は史上最大の5230万台に達したという。

株式会社I&M研究所  
2008年1月29日、東京、午後、午後2時30分、07年度国内携帯電話台数発表会（07年3月30日）の発表結果を発表。発表結果は、07年度末では、携帯電話の総台数は、国内最大の5230万台に達した。発表結果は、国内最大の携帯電話台数を記録した2007年を上回り、日本の携帯電話台数は史上最大の5230万台に達したという。

Japan Internet.com Webビジネス - 国内最大の携帯電話台数を記録——シェアトップは、ガートナー・ジャパンは、2007年の国内における携帯電話の総台数が、5230万台に達したと発表した。カマクラ・ジャパンの発表で、国内最大の携帯電話台数を記録した2007年を上回り、日本の携帯電話台数は史上最大の5230万台に達したという。

国内の携帯電話台数は、国内最大の5230万台に達した。発表結果は、国内最大の携帯電話台数を記録した2007年を上回り、日本の携帯電話台数は史上最大の5230万台に達したという。

主観の携帯電話台数は、国内最大の5230万台に達した。発表結果は、国内最大の携帯電話台数を記録した2007年を上回り、日本の携帯電話台数は史上最大の5230万台に達したという。

契約数

携帯電話/PHS接続サービス(携帯)/PHS/無線呼出し契約数

事業年度	1	2	3	4	5	6	7	8	9	10	11	12
平成25年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成24年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成23年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成22年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成21年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成20年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成19年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成18年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成17年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成16年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成15年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成14年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成13年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成12年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成11年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成10年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成9年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成8年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2
平成7年	11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0	12.1	12.2



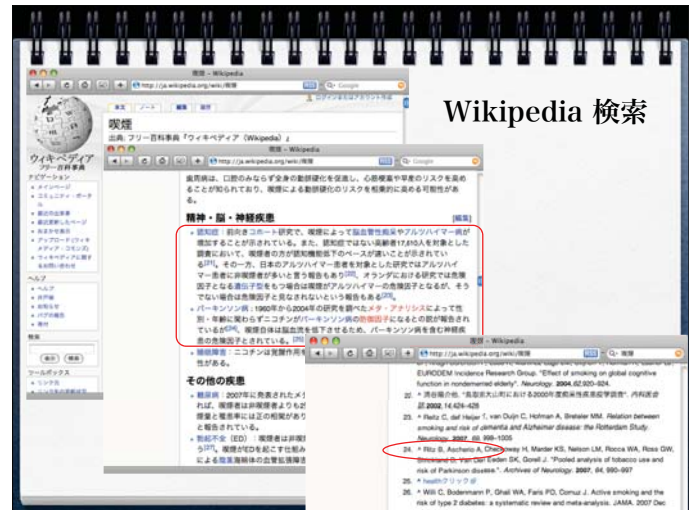
## 検索のポイント 1

- 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。

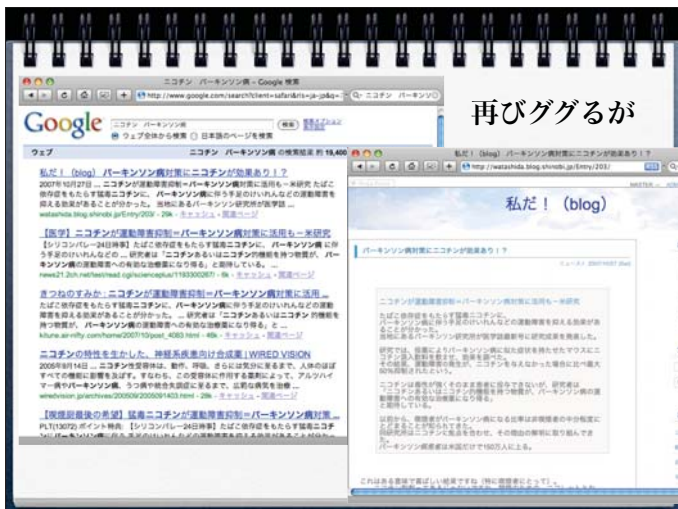
タバコを吸っているとある病気に罹りづらいという噂を聞いた。その病気とは何か。



## Wikipedia 検索

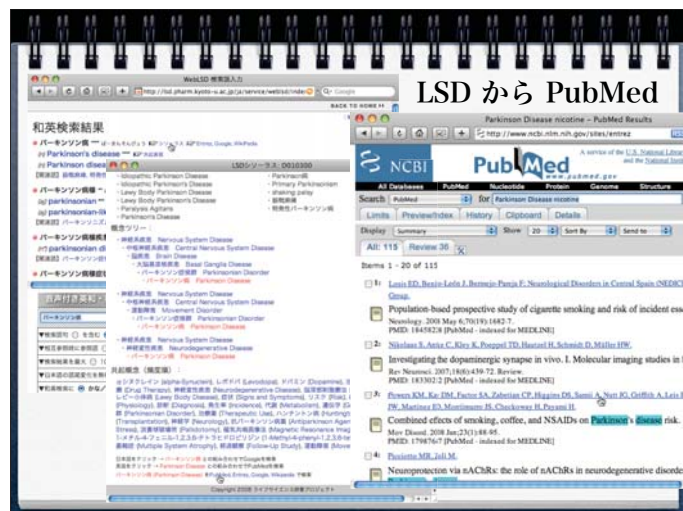
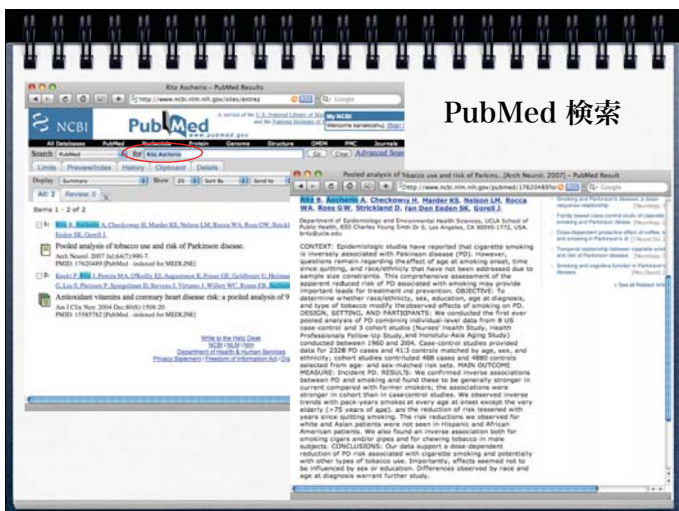
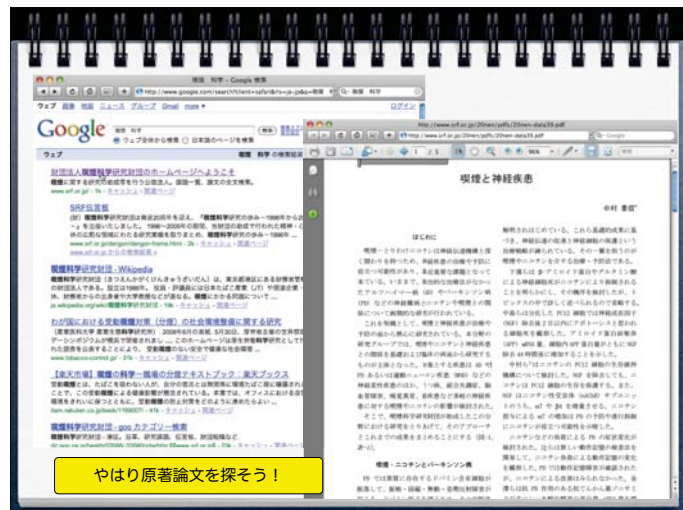
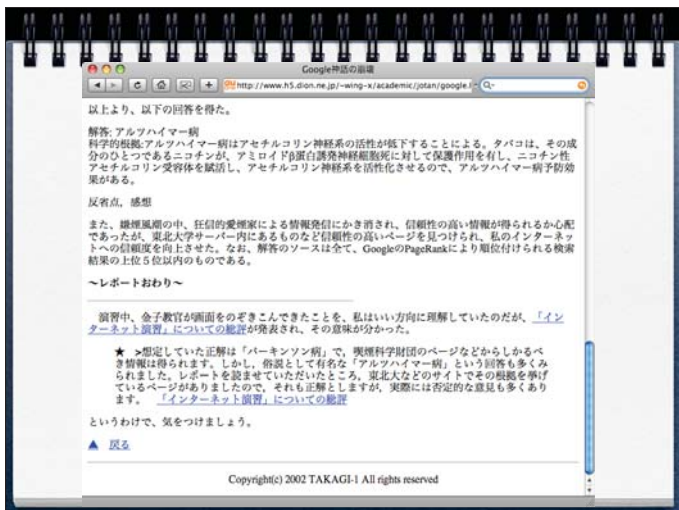
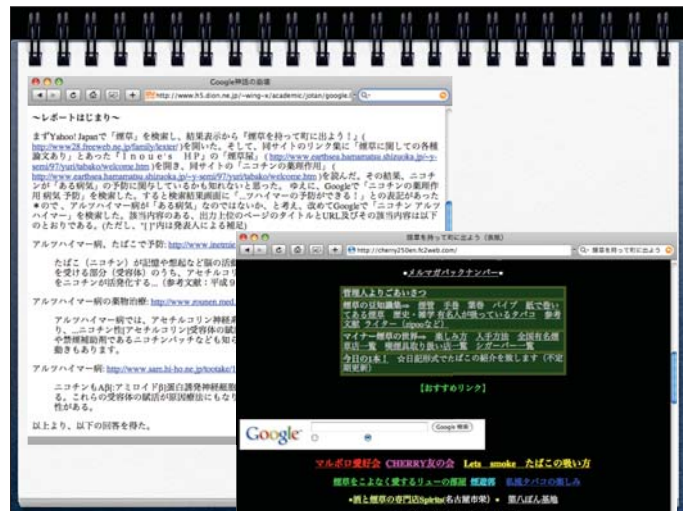
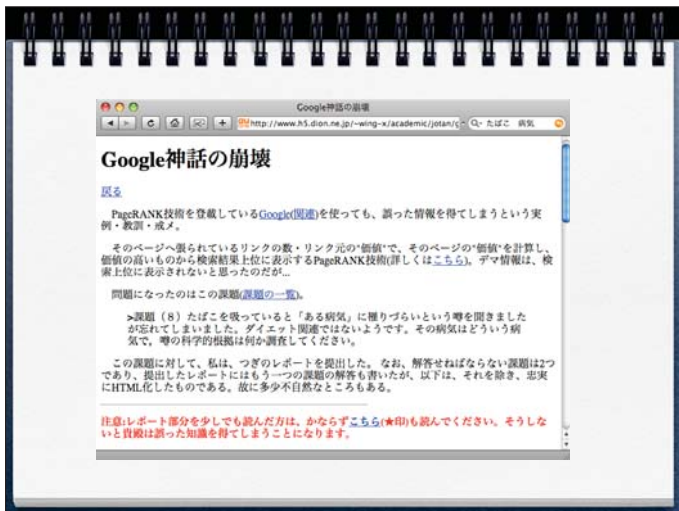


## 再びググるが



## あつ





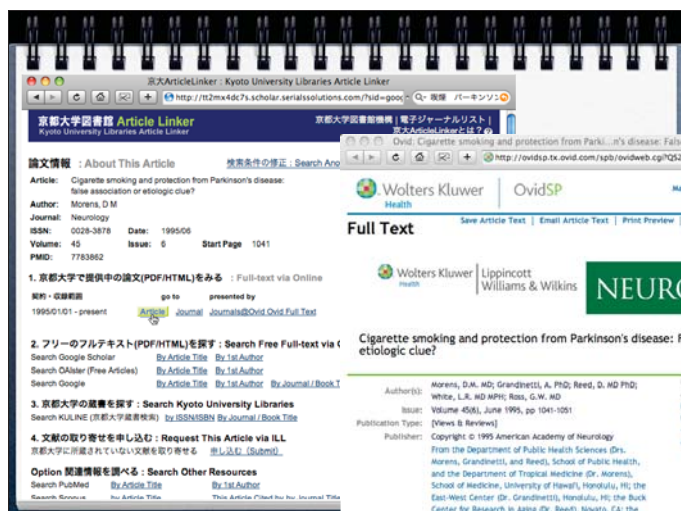




## Web of Science 検索



## Google Scholar 検索



## 英語の医学系論文を探す

- 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。

## 検索のポイント 2

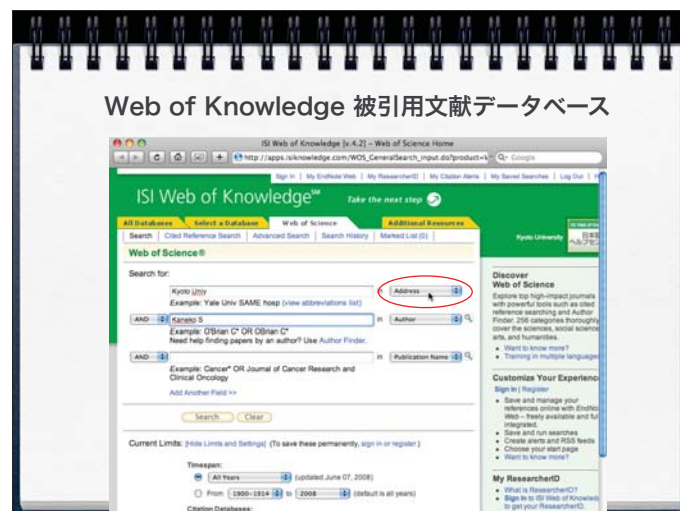
- 表記のゆれ、同義語、言い換え表現
- ネットには諸説入り乱れている
- 科学的な情報＝総説や論文を探す
- 科学的情報は英語で検索する
- 科学的情報に「定説」はない

## データベース演習

- 課題：京大に在籍する 1 名の教員について、専門研究の業績（学位論文、著作物や論文と被引用状況）や 科研費取得状況を調査する。
- 目標
  - 代表的データベースを使えるようになる。
  - 情報を要約することができる。
  - 情報がない原因を考えられる。



出所: <http://read.jst.go.jp/>







## 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