

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

大学図書館の活用と情報探索 2017年5月30日

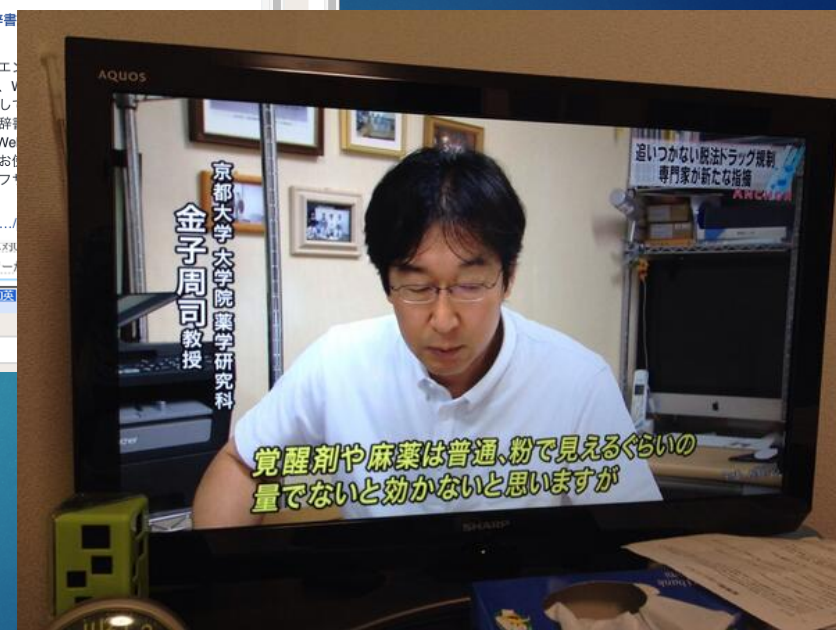
薬学研究科 金子 周司

自己紹介



2

2. 医学辞書づくり (計量言語学で)



1. 薬理学 (痛みの仕組みを調べて、新しい鎮痛薬をつくる研究)

3. 麻薬の専門家

研究における情報の重要性

3

- ▶ 大学とは、未知のことを自分で研究するところ（卒論、D論）
- ▶ どうやって、研究テーマを見つける？
 - ▶ すでに先人が十分に研究してしまっていたら、二番煎じに...
- ▶ 研究成果の価値は「新しい」こと
- ▶ 先行研究を調べ上げて新しいテーマを探すのは、なかなか難しい
- ▶ 図書だけでなく、インターネットやデータベースもフル活用したい

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（何が既知で何が新しいのか？）
 - ▶ そもそも科学においては、既存の知識すら疑うべき
- ▶ 研究や調査をする技術（スキル）
- ▶ 研究コミュニティ（人脈）：「情報を発信する人に情報は集まる」
- ▶ 語学力、プレゼン能力
 - ▶ もちろん資金も...

情報の検索にあたってのスキル

- ▶ 皆さんはネット検索なら慣れている？
- ▶ 一般の情報検索は、網羅的でなくても多くは問題ない
 - ▶ 例えばレストラン探しとか、YouTubeでのビデオ探しとか...
- ▶ でも、先行研究の情報検索は、見落としがあるのはマズイ
- ▶ どうしたら、十分かつ最新の情報を探し出すことができる？
- ▶ 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



書籍
(電子書籍)

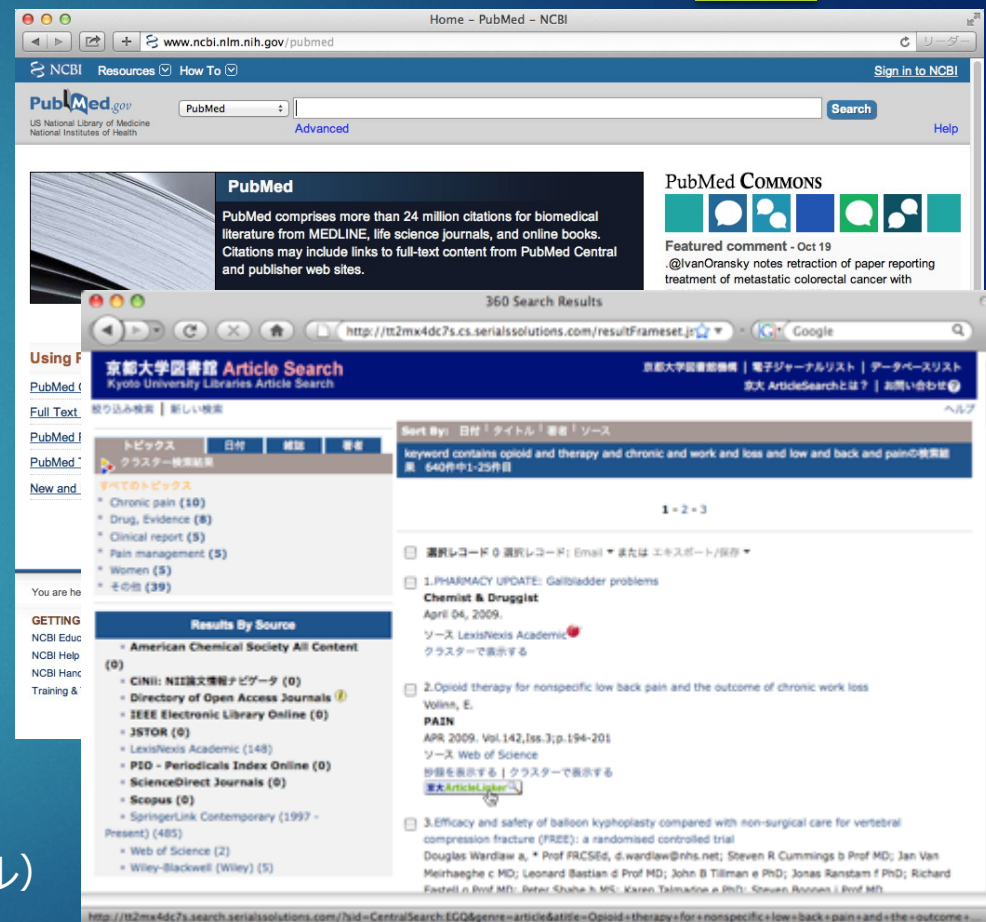


学術誌 (冊子体、電子ジャーナル)



文献抄録データベース

6



電子図書館

資料の特性

Peer Review

A process by which a scholarly work is checked by a group of experts in the same field to make sure it meets the necessary standards before it is published or accepted

7

- ▶ 書籍、学術雑誌
 - ▶ 一次情報、査読された内容、最近どんどん電子化、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

- ▶ 世界中に流通させられる
- ▶ 早く読者に届けられる
- ▶ 流通が少数でも、出版のようにコストがかからない
- ▶ 書かれる論文数が膨大であっても、収蔵にスペースが必要ない
- ▶ 検索できないと意味が無いが、索引づけが半自動的である
- ▶ 引用や参照をリンクとして実装できる

学術誌のステイタス

9

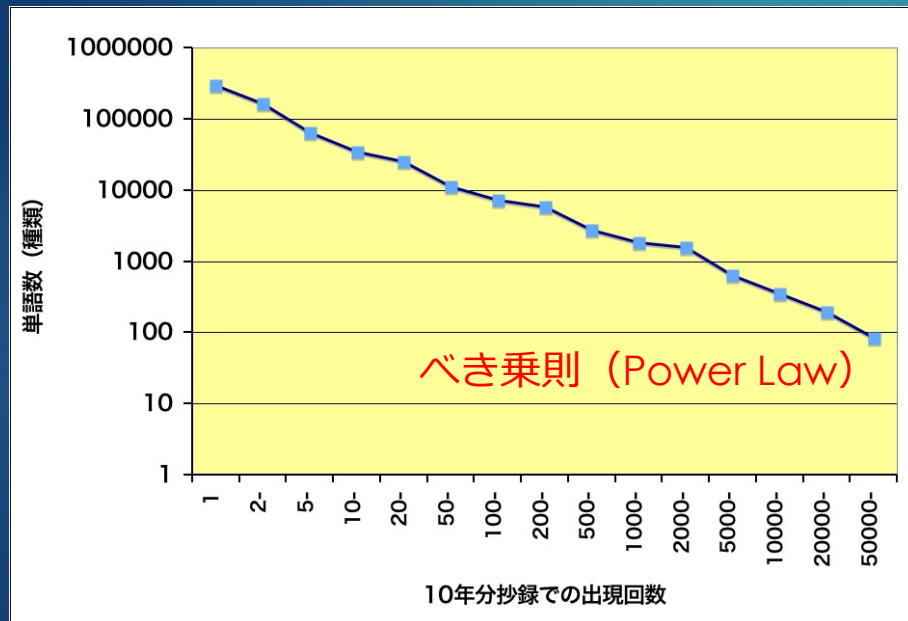
最近では、個別の論文について被引用回数やSNSでの評判などから影響度を測ろうとする動きも盛ん

- ▶ Impact Factor (IF)
 - ▶ 論文ごとに引用リンクがある仕組みを応用している
 - ▶ よく引用されている論文ほど、学界にインパクトを与えている
 - ▶ インパクトの高い論文を多く掲載している学術誌ほど、ステイタスがある
 - ▶ 学術誌ごとに次式でインパクトファクターが毎年求められる
 - ▶ ある1年での前々年と前年の論文の総被引用数 ÷ 前々年と前年の掲載論文数
= 1論文あたり発表後2年間での被引用数の平均値
 - ▶ Journal of Citation Reports にて調べられる

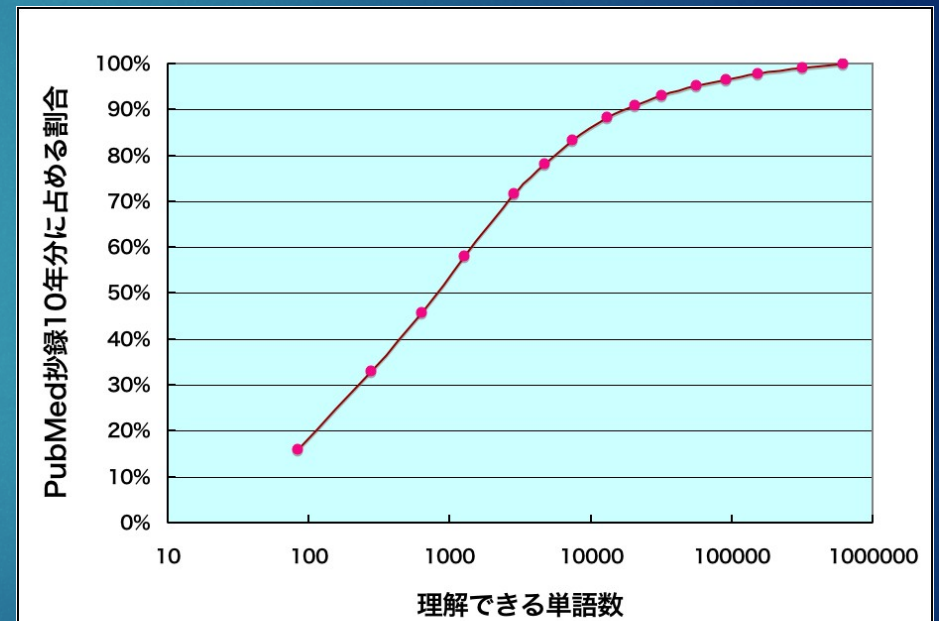
ライフサイエンス辞書

10

- ▶ 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



60万語のうち30万語は1回しか出現しない
(全体の1%未満)



頻度上位10万語で97%は解読できる

Key Word in Context (KWIC)

11

The screenshot displays the 'LIFE SCIENCE DICTIONARY' website. The search term '考え' (thought) is entered in the search box. The results are sorted by '1語後でソート' (Sort by 1 word after). The left sidebar shows the search results for '考え' and related terms like 'idea', 'notion', and 'opinion'. The main content area shows a list of search results with KWIC examples. The examples are numbered 1 through 37 and show the word 'thoughts' in context. The examples are:

- 1 their loved one (grief state) and a neutral
- 2 ted whether having students write down their
- 3 l assessment of lymph nodes, and the current
- 4 (OR, 2.0; 95% CI, 1.1 to 3.7); and avoiding
- 5 he training institution, residents' personal
- 6 lyzed: thoughts about death, wanting to die,
- 7 ideation resolves rapidly, the resolution of
- 8 Diagnostic Interview Schedule were analyzed:
- 9 95% confidence interval [CI], 1.2 - 4.6) for
- 10 l urges and improves cognitive resistance to
- 11 a quarter of oncology patients had seriously
- 12 First, we experimentally induced
- 13 ues described in the above papers along with
- 14 The bonding environment of boron is usually
- 15 ast pain, mammography pain, and catastrophic
- 16 g and Antithrombotic Regimen have stimulated
- 17 t has been claimed that this is the case for
- 18 ody schema, autonomic response, catastrophic
- 19 For instance, deepening our
- 20 portant to try to ascertain what the patient
- 21 ve emotional reactions to smoking and induce
- 22 49), 3 (n = 52), and 4 (n = 43) manipulated
- 23 Here, we discuss some worrisome
- 24 We conclude with
- 25 of the current achievements, and offer some
- 26 es in self-esteem, body image, and intrusive
- 27 ndings suggest a need to re-evaluate current
- 28 nce in this environment, I also include some
- 29 We conclude with some
- 30 ections of the author and ends with a few
- 31 Dysfunctional negative
- 32 ternated between recalling a sad, ruminative
- 33 We have asked three individuals who have
- 34 They close with some
- 35 ient described the control as intuitive; she
- 36 ous recollections of these meetings and some
- 37 elop insight concerning the general public's

The right sidebar shows the search results for '考え' and related terms like 'idea', 'notion', and 'opinion'. The results are sorted by '1語後でソート' (Sort by 1 word after). The examples are numbered 1 through 37 and show the word 'thoughts' in context. The examples are:

- 1 thought about a different person they knew an equally lo
- 2 thoughts about an upcoming test could improve test perfo
- 3 thoughts about best surgical practice in the management
- 4 thoughts about breast cancer (modest: OR, 2.2; 95% CI, 1
- 5 thoughts about breast imaging, and their interest in per
- 6 thoughts about committing suicide, and attempted suicide
- 7 thoughts about death is more gradual.
- 8 thoughts about death, wanting to die, thoughts about com
- 9 thoughts about death; 4.4 (95% CI, 1.7 - 11.6) for wante
- 10 thoughts about drinking.
- 11 thought about euthanasia or physician-assisted suicide a
- 12 thoughts about finances and found that this reduces cogn
- 13 thoughts about future progress of the zebrafish model as
- 14 thought about in terms of localized 2c-2e/3c-2e bonding
- 15 thoughts about mammography pain.
- 16 thoughts about new adjunctive agents and devices, includ
- 17 thoughts about numbers: Children cannot have the concept
- 18 thoughts about pain, and fear of movement (r > 0.42, P <
- 19 thoughts about personhood should support the development
- 20 thought about quality-of-life values before surrogate de
- 21 thoughts about quitting; however, contrary to models of
- 22 thoughts about science and examined the causal impact of
- 23 thoughts about the condition.
- 24 thoughts about the current state of antimalarial chemoth
- 25 thoughts about the future of this field.
- 26 thoughts about the illness.
- 27 thoughts about the immune suppression that might occur d
- 28 thoughts about the opportunities and challenges of under
- 29 thoughts about the potential therapeutic efficacies of H
- 30 thoughts about the present status and future prospects o
- 31 thoughts about the self have long been hypothesized to r
- 32 thought about their loved one (grief state) and a neutra
- 33 thought about this very complex issue to share their rat
- 34 thoughts about training future epidemiologists to best t
- 35 thought about using her hand or elbow and the prosthesis
- 36 thoughts about why they were of such central importance
- 37 thoughts about witnessed CPR.

事物を表す言葉には「ゆれ」がある

12

語	総説誌A	公定書B
抗癌薬	19	0
抗癌剤	763	32
抗がん薬	0	9
抗がん剤	3	6
制癌剤	37	3
抗腫瘍薬	13	0
抗腫瘍剤	12	52
抗悪性腫瘍薬	10	13
抗悪性腫瘍剤	1	741
悪性腫瘍治療薬	1	0

- ▶ 英語でも似たような状況
 - ▶ Agent, Drug...
 - ▶ Cancer, Tumor, Neoplasm...
- ▶ 検索語の単純一致では、多くの情報を見逃すことに
- ▶ 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
- ▶ しかし、全ての情報検索システムが「賢い」とは思わないこと

概念や事物を表すことば

13

- ▶ 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である
 - ▶ 「京都」では範囲が不明瞭である → 正しく限定する必要 = 「粒度」
- ▶ ことばには「ゆれ」（同義語、異表記、略語など）がある
 - ▶ たんぱく質、蛋白質、タンパク質、プロテイン...
- ▶ 人物の場合の同姓同名など、固有名詞の重複もよくある
- ▶ 日本語だけの調査は不十分である

シソーラスthesaurus (同義語辞書)

14

- ▶ 概念を表す統制語 descriptor を決め、同義語 synonym を寄せ集める
- ▶ 概念の上位下位によりツリーを作る
- ▶ 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

The screenshot displays the LIFE SCIENCE DICTIONARY interface. The top navigation bar includes 'WebLSD', '英語教材', '変換サービス', 'アプリ・本', and 'プロジェクト'. Below this, there are tabs for '英和・和英', 'シソーラス', and '英語共起表現'. The search bar contains 'Vitamin D' and 'Diabetes Mellitus'. The results for 'Vitamin D' show synonyms like 'ビタミンD' and a concept tree including 'Polycyclic Compound', 'Steroid', and 'Secosteroid'. The results for 'Diabetes Mellitus' show synonyms like '糖尿病' and a concept tree including 'Nutritional and Metabolic Disease', 'Metabolic Disease', and 'Glucose Metabolism Disorder'.

検索スキルを上げるために

- ▶ 情報源によって内容が違うので、複数の情報源を使いこなす
- ▶ 索引システムが異なるので、検索語を様々に変えて探索する
- ▶ いつ、誰が、どこに残した情報か、信憑性や客観性を確かめる
- ▶ 複数キーワードや演算子によつて的確に絞り込む
- ▶ 重要な情報は何か考えて整理する（被引用調査の必要性）
- ▶ 重要な情報は、一次情報にまで遡って調べる（情報源はどこか）
- ▶ 情報が得られないなら、理由を考えて見逃しを防ぐ

この単元の演習

16

- ▶ 総合演習と関連させる
- ▶ まず今からグループに分かれて、調査するテーマを話し合って決める
- ▶ 来週のデータベース演習では、そのテーマについて詳しい専門家の大学教授1人（できれば京大の中で）について、業績などを調べる

来週する演習の目的

- ▶ 下記データベースを利用して研究業績を調査せよ。
- ▶ 京都大学教育研究活動データベース
- ▶ 科学研究費助成事業データベース KAKEN
- ▶ 国立情報学研究所 CiNii Articles
- ▶ 科学技術振興機構 J-GLOBAL
- ▶ トムソン・ロイター Web of Science
- ▶ いくつかの新聞データベース（朝日、毎日、読売）

レポートに記載する内容

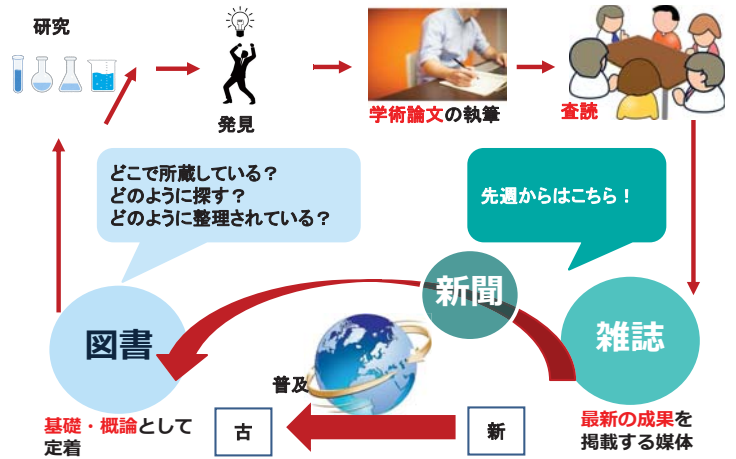
- ▶ 氏名、現職
- ▶ 研究開始年がもっとも新しい科学研究費補助金の獲得状況
- ▶ 2006年以降の論文で最も被引用数の多いものの書誌情報
(日本語、英語 計2件) およびその被引用数
 - ▶ 共著者もすべて解答すること。
- ▶ 教員に関する2006年以降の新聞記事
- ▶ それらの結果を踏まえた考察

演習レポートの考察

19

- ▶ 同性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。
- ▶ それぞれのデータベースの特徴を比較してまとめよ。
- ▶ 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。
- ▶ 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。

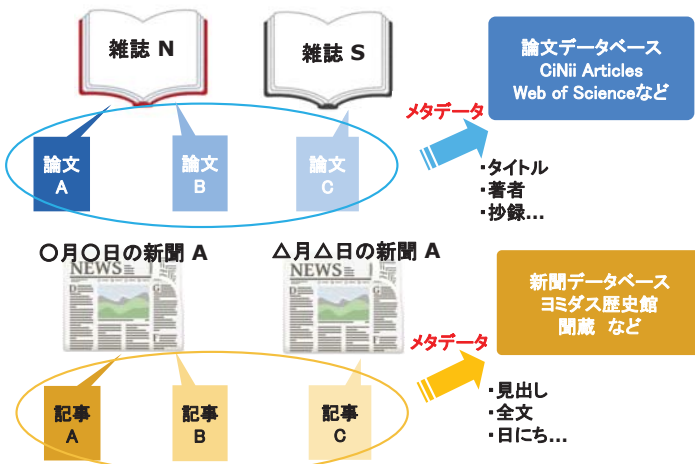
おさらい 学術情報の生産



演習のポイント

2017/6/6 大学図書館の活用と情報探索
「インターネット検索の落とし穴とデータベースの活用法」

おさらい データベースとは



データベースとは

- ・特定のテーマに沿ったデータを
- ・記録、収集し
- ・容易に検索、抽出できるようにしたもの

そのため
 ・収録内容 ・収集範囲
 ・検索方法 ・インターフェース
 などは作成者(DB)によって異なる!

Q 課題で使うデータベースにはどうやってアクセスするの?

- A1 PandAの課題ページ下部にリンク先をつけてあります
- A2 京都大学データベースリストを活用しよう



お願い! 「ログアウト」を忘れずに

Web of Science 検索の前に

データベースリストからアクセスした場合
横断検索を
Web of Science Core Collectionに
切り替えましょう！



<https://apps.webofknowledge.com/>

Q Web of Scienceで「著者」での検索が できません

A Web of Scienceの検索のルールに従っていますか？
→著者名は名字、スペース、名前の順に入力(ガイド p.1右下)



データベースによって検索ルールが異なる！

大文字と小文字は
区別するの？

AND, OR検索の方法は？

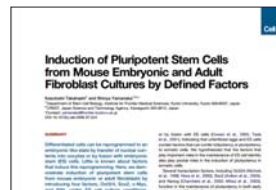
よくわからないけど
上手いかない・・・

各データベースの「ヘルプ」をチェック！

→課題ページ下部にまとめています

補助者にお気軽に質問してください！

Q 被引用件数とは何ですか？



Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-676. doi:10.1016/j.cell.2006.07.024



被引用件数 =

ある論文が、他の論文内で引用された数

例えば...

2006年に「CELL」誌に掲載された
山中伸弥先生の論文
「Induction of pluripotent stem cells
from mouse embryonic and adult
fibroblast cultures by defined factors」
の被引用件数を調べる

なぜ違いが出てくるのか

データベース	被引用数 2017/6/1 現在	(参考) 被引用数 2014/11/17 現在
Web of Science	9986	6742
CiNii Articles	157	157
Google Scholar	17122	10941



各データベースの収録範囲が異なるから

Q 新聞記事を調べるデータベースとは？

A それぞれの特徴を意識しながら検索してみてください

	キーワード 検索	全文検索	主要雑誌 タイトル	用語辞 典
聞蔵II [キクゾウツー] (朝日新聞)	1879～	1985～	週刊朝日(2000～) AERA(1988～) アサヒグラフ(1923- 1956)	○
ヨミダス 歴史館 (読売新聞)	1874～	1986～	-	○
毎索 [マイサク] (毎日新聞)	1872～	1987～	週刊エコノミスト (1989～)	×

まとめ

- DBを利用したらログアウトをお願いします
- 検索記号や様々な条件を上手に使おう
- ヘルプを参照しながら、試行錯誤しつつ
いろいろなデータベースを活用してみてください
- 困ったらお気軽に補助者までお声掛けください！

参考資料

①代表的な論文データベースの紹介

②論理演算の入力方法の紹介

- CiNii Articles
- 新聞データベース
- Google

①代表的な論文DB(日本語)

● CiNii Articles [サイニイ アーティクルズ][全般]

<http://ci.nii.ac.jp/>

国立情報学研究所(NII)が提供している、日本の**学術論文**を中心とした論文情報の提供サービス。

● JDream III [ジェイドリーム・スリー][全般/科学技術]

<http://jdream3.com/>

株式会社ジー・サーチ提供の**科学技術に関する文献や研究テーマ情報**などを検索することができるデータベース。科学技術系のジャーナルを初め、学会誌、協会誌、企業・大学・独立行政法人・公設試験場等の技術報告、業界誌、臨床報告等を収録。

● 医中誌Web [イチユウシ ウェブ](1977-)【医学】

<http://search.jamas.or.jp/>

医学中央雑誌刊行会が作成・提供する**国内医学論文情報**のインターネット検索サービス。医学・歯学・薬学および看護学・獣医学などの関連分野の定期刊行物、のべ約6,500誌から約1100万件(2017年6月1日現在)の論文情報を収録。

①代表的な論文DB(英語)

● Web of Science [ウェブ オブ サイエンス][全般/科学技術]

<http://webofscience.com/>

トムソン・ロイター社提供。**自然科学、社会科学、人文科学**のジャーナル12,000誌以上が検索でき、引用情報も調べられる。データの収録範囲は**1900年**までさかのぼる。

● Scopus [スコパス][全般/科学技術]

<https://www.scopus.com/>

エルゼビア社提供。**自然科学、社会科学、人文科学**のジャーナル21,000誌以上が検索できる、世界最大規模のデータベース。抄録は最も古いものは1800年代までさかのぼり、1996年以降の論文は引用情報も収録。

● PubMed [パブメド][医学]

<https://www.ncbi.nlm.nih.gov/pubmed?otool=ijpctlolib> (京大専用)

米国国立医学図書館(NLM)提供【無料】

医学・薬学・生物学分野の1946年以降の論文を収録。

②論理演算の入力方法-CiNii Articlesの場合

	入力方法	意味
AND検索	情報処理_デジタル スペース	「情報処理」「デジタル」両方を含む
OR検索	原住民 OR 先住民 半角大文字	「原住民」「先住民」どちらかを含む
NOT検索	福祉 -ボランティア 半角ハイフン	「福祉」を含むが「ボランティア」を含まない
完全一致検索	/清水光/	「清水光ー」はヒットしない
前方一致検索	カウンセ*	カウンセラー、カウンセリング...
フレーズ検索	"Google Maps"	2語が隣り合うものだけ

②論理演算の入力方法-新聞DBの場合

	意味	入力例【ヨミ・毎索】※	入力例【間蔵】※※
AND検索	○と△の両方を含む	○_AND_△ ○スペース△	○&△ ○スペース△
OR検索	○または△を含む	○_OR_△	○+△
NOT検索	○は含むが△は含まない	○_NOT_△ ○_ - △※※※	○#△
検索の順序付け	()内を優先して検索	(○_OR_△)_NOT_□	(○+△)#□

※半/全、大/小問わず入力可。英語/記号の前後にスペースを入力。

※※ &、+、#、()は半角で入力。優先順位はNOT>AND>OR

※※※ 「-」記号での検索はヨミダスのみ利用可。

②論理演算の入力方法-Googleの場合

https://www.google.co.jp/

Google
日本

環境社会 技術 エネルギー

Google 検索 I'm Feeling Lucky

(Googleの場合)

- 論理演算

- ✓ AND検索 A B
- ✓ OR検索 A OR B、A|B
- ✓ NOT検索 A -B

フレーズ検索 “All roads lead to Rome”

