

生命科学情報の統合に向けて

情報検索技術について

大学共同利用機関法人
情報・システム研究機構
ライフサイエンス統合データベースセンター
川本祥子

世界に生命科学関係のデータベースは どれくらいあるでしょう



Nucleic Acids Research Database issue

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

[Oxford Journals](#) > [Life Sciences](#) > Nucleic Acids Research

READ THIS JOURNAL

Nucleic Acids Research
Volume 38, Issue 1, January 2010

[View Current Issue \(Volume 38 Issue 1 January 2010\)](#)
[Advance Access](#)
[Browse the Archive](#)

New! 2010 Database Issue and Database Collection. The latest Database issue includes descriptions of 58 new and 73 updated data resources. The accompanying online Database Collection lists 1230 carefully selected databases covering various aspects of molecular and cell biology. [Read more](#)

NAR's Featured Articles represent the top 5% of papers in terms of originality, significance and scientific excellence. Check these [here](#)

Nucleic Acids Research is proud to have been chosen by the Special Libraries Association as [one of the top 100 most influential journals in medicine and biology](#).

SEARCH THIS JOURNAL > [Advanced search](#)

CLICK HERE TO READ NAR'S 2010 DATABASE ISSUE

RELATED JOURNALS

- > [Bioinformatics](#)
- > [Database](#)
- > [Human Molecular Genetics](#)
- > [Glycobiology](#)
- > [Mutanagenesis](#)

READER SERVICES

- > [Careers Online](#)
- > [Calendar of Events](#)
- > [Societies in the Nucleic Acids Area](#)
- > [50 Most Frequently Cited Articles](#)
- > [NAR Institutional Membership](#)

THE JOURNAL

- [About this journal](#)
- > [NAR Methods online](#)
- > [2010 Database Issue](#)
- > [2009 Web Server Issue](#)
- > [Nucleic Acids Research](#)
- > [Referee Information](#)
- > [Rights & Permissions](#)
- > [Dispatch date of the next issue](#)
- > [This journal is a member of the Committee on Publication Ethics \(COPE\)](#)

Impact factor: 6.878

Senior Executive Editors
Keith Fox, Southampton, UK
Barry Stoddard, Seattle, WA, USA
> [View full editorial board](#)

FOR AUTHORS

- > [Instructions to authors](#)
- > [Scope and Criteria for Consideration](#)
- > [Online submission instructions](#)
- > [Submit a manuscript now](#)

•1993年、24のデータベースに関する論文から始まったデータベース特集号

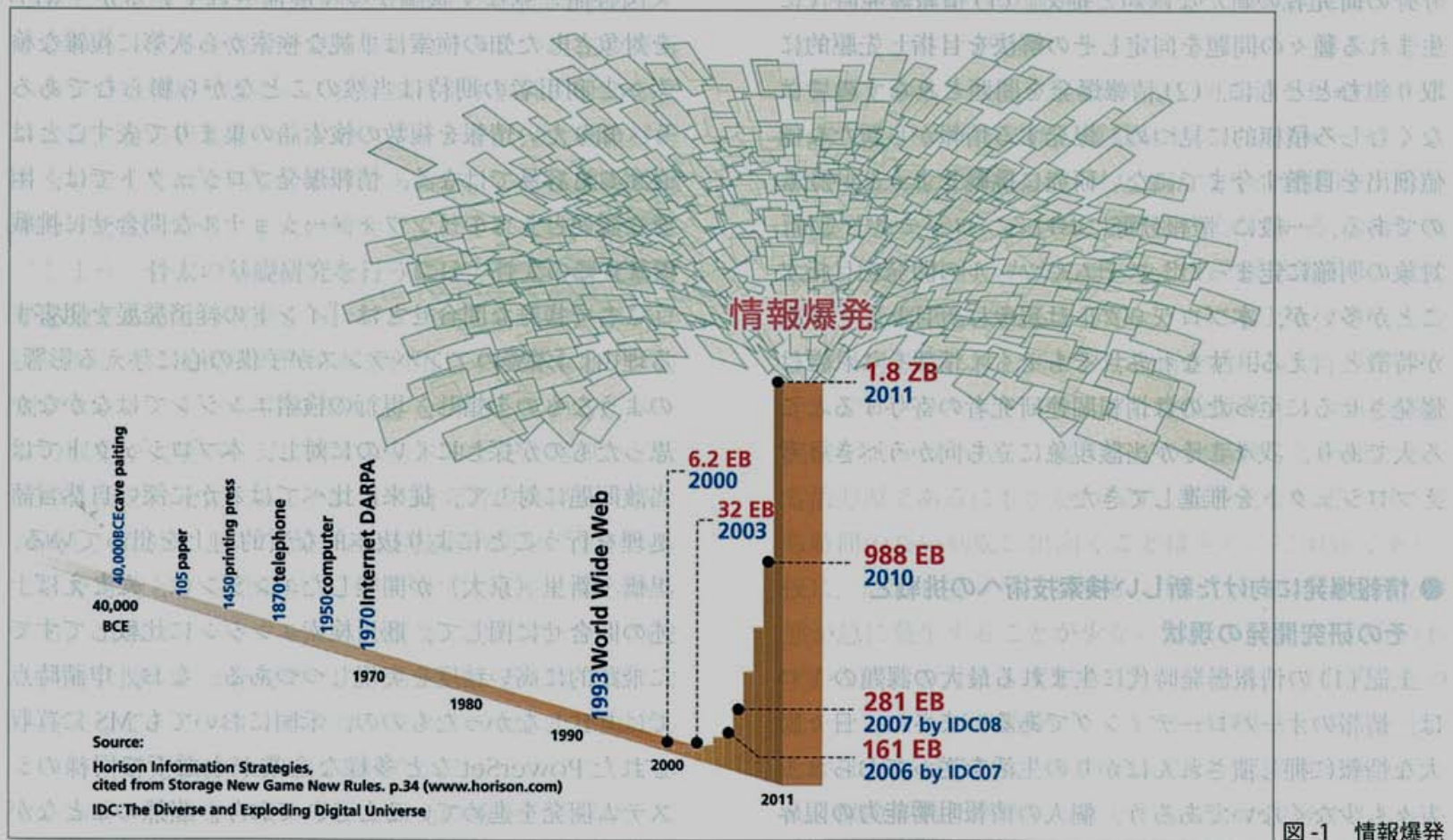
•2010年度版で
合計1230データベースに

•オープンアクセスで全部読めます！

•<http://nar.oxfordjournals.org/>

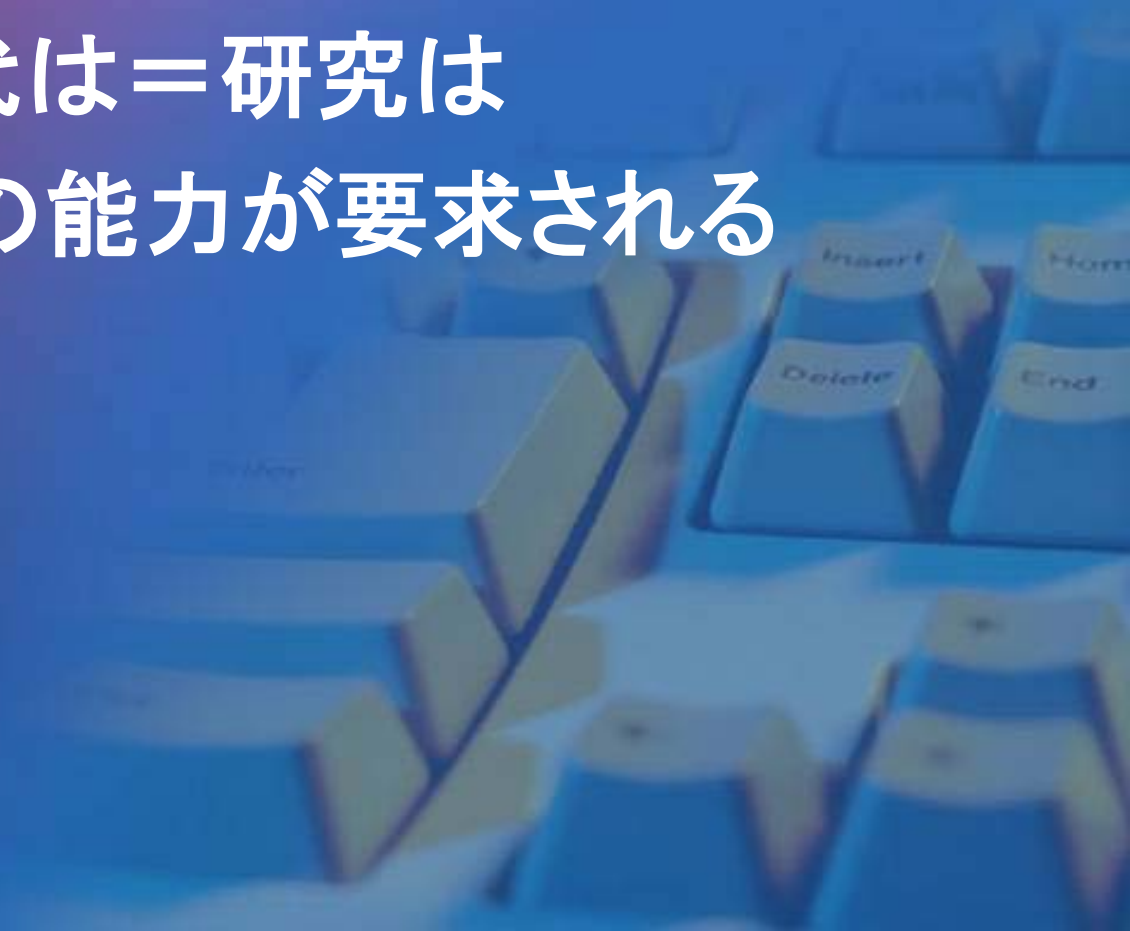
データが多いのは生命科学に限らない問題

情報爆発時代： 人類が有史以来作製した文書数の推移



☆¹ IDC の 2007 年予測値は 2008 年に修正されたが、本図は、両者とさらに 2004 年特定申請当時 Horizon 社から公開された数字を合わせて描いた。

現代は＝研究は
情報検索の能力が要求される



世界中の情報を検索できる便利な時代

Google 久留米 ラーメン 検索 検索オプション

☒ ウェブ全体から検索 ☐ 日本語のページを検索

ウェブ 検索ツールを表示 久留米 ラーメン の

他のキーワード: 久留米ラーメン ランキング 久留米ラーメン 太砲 久留米ラーメン 通販 久留米ラーメン 清陽軒 久留米ラーメン

久留米駅(福岡) 付近のラーメンの検索結果



A 大茶ラーメン
daieiramen.fc2web.com - 0942-39-3977 - レビュー (8)

B 丸幸ラーメンセンター原古賀店
www.marukoura-men.com - 0942-40-3737 - レビュー (7)

C 大龍ラーメン ベルモール東町店
www.dairyuramen.com - 0942-36-0446 - レビュー (4)

D 来福軒
maps.google.co.jp - 0942-33-7957 - レビュー (2)

E らーめん亭
maps.google.co.jp - 0942-35-7472 - レビュー (6)

F 太砲ラーメン 長門石店
www.taiho.net - 0942-39-7424 - レビュー (3)

G 清陽軒本店
maps.google.co.jp - 0942-35-2237 - レビュー (5)

久留米駅(福岡) 付近のその他の検索結果 >

久留米ラーメン - Wikipedia

久留米ラーメン(くるめラーメン)は、福岡県久留米市を中心に作られる、豚骨スープとストレートの細麺をベースにしたラーメンである。豚骨ラーメンの元祖ともいわれているラーメンで、多くの九州ラーメンだけでなく、様々な地域のラーメンにも影響を ...

[発祥](#) - [特徴](#) - [影響](#) - [関連項目](#)

ja.wikipedia.org/wiki/久留米ラーメン - キャッシュ - 類似ページ - 印刷

久留米 太砲ラーメン - [イントロをスキップ]

呼び戻しスープ発祥の老舗。とんこつラーメンの神髄。

www.taiho.net/ - キャッシュ - 類似ページ - 印刷

ちょっと待った

Googleの処理対象URL 1兆突破

日本のブログ数 1690万

記事総数13億

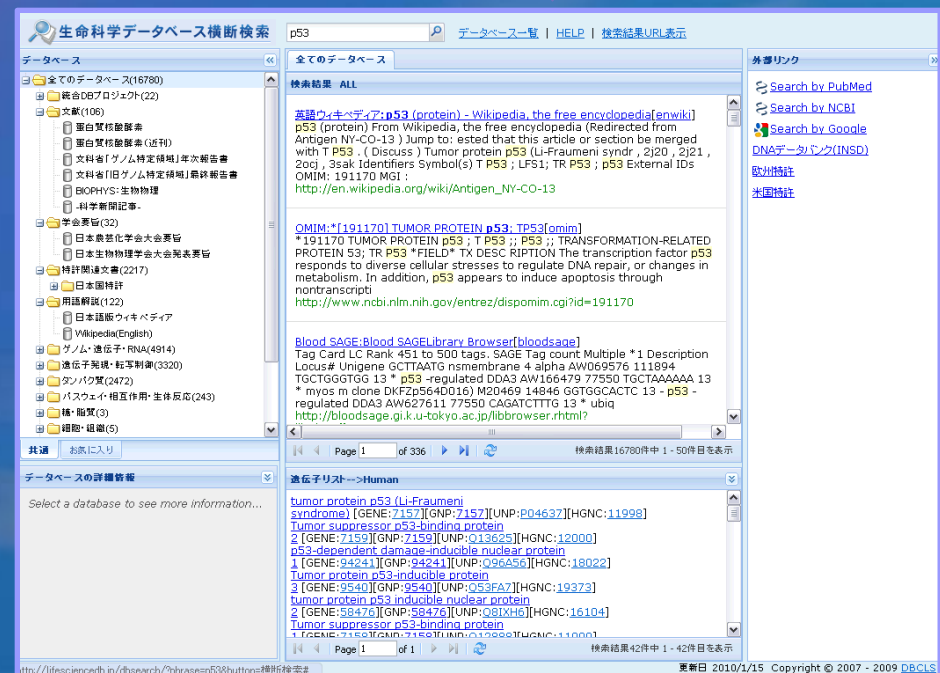
Pubmed 1859万

Genbank 核酸 7500万

(2008年データ)

統合DBプロジェクトでは 生命科学に特化した検索サービスを提供

- 生命科学データベース横断検索 } 200の生命科学DB
- OReFiL
- Allie } Pubmed
- inMeXes



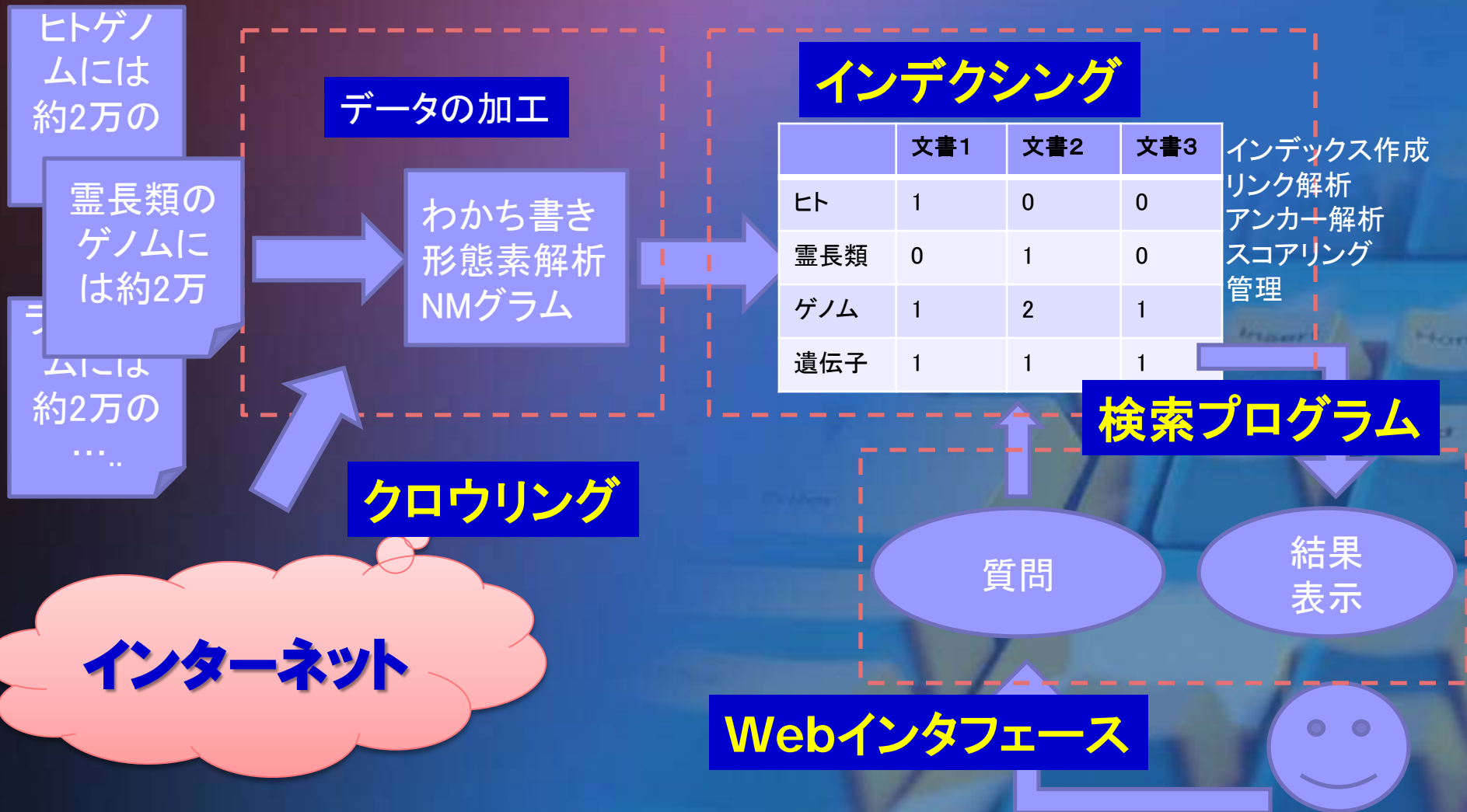
検索エンジンとは

- 情報の収集(クロウリング)、インデクシング、検索という3要素で成り立つシステム
- 大きく分けて二つのタイプ
 - ディレクトリ型--- 昔のYahoo, 統合DBのDBカタログ
 - ロボット型--- Google, 統合DBの横断検索

文章の中からキーワードを検索

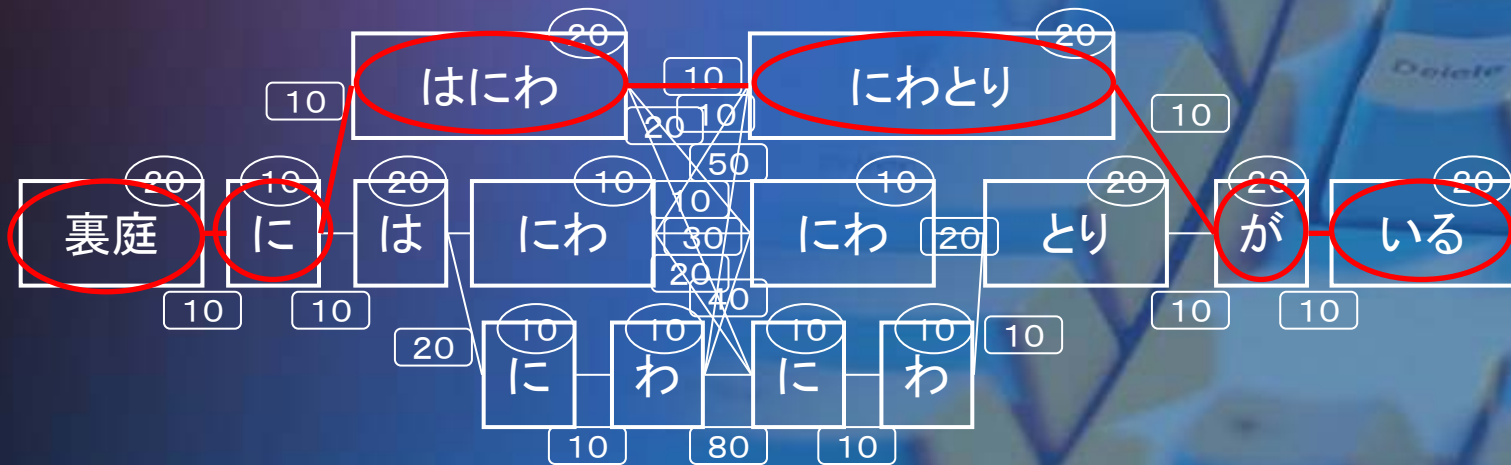
- 1, 特定領域研究「生命システム情報」及び特定領域研究「比較ゲノム」と連携し、ゲノム研究関連特定研究領域で解析の対象となる真核生物であるメダカ(近縁種2種)、ホヤ(各種)、ナメクジウオ、ギボシムシ、**霊長類**(各種)、カイコ、コムギ、クラミドモナス等を中心に、オリゴキャップ法による完全長cDNA ライブラリー、5' 端濃縮ライブラリー、5' 端SAGE(5' SAGE)ライブラリーを作製し、それらのcDNA ライブラリーからのクローンの5' 端EST 配列を決定することで、遺伝子の転写開始点の詳細とそのコアプロモータ領域を同定する。

検索の基本的な仕組み



(辞書引きアルゴリズム)

- 裏庭にはにわにわとりがいる



配列検索

```
AGCTGTCTGTAGCTAGCTAGCAT  
CCTAAACCCTGTGTGGGGGCATG  
AGCTGTCTGTAGCTAGCTAGCAT  
CCCTGTGTGGGAATAAGGCAT  
CTTTTGTATGTGTGTGTGTCCCC  
CTTTTGTATGTTAGTCCCC
```

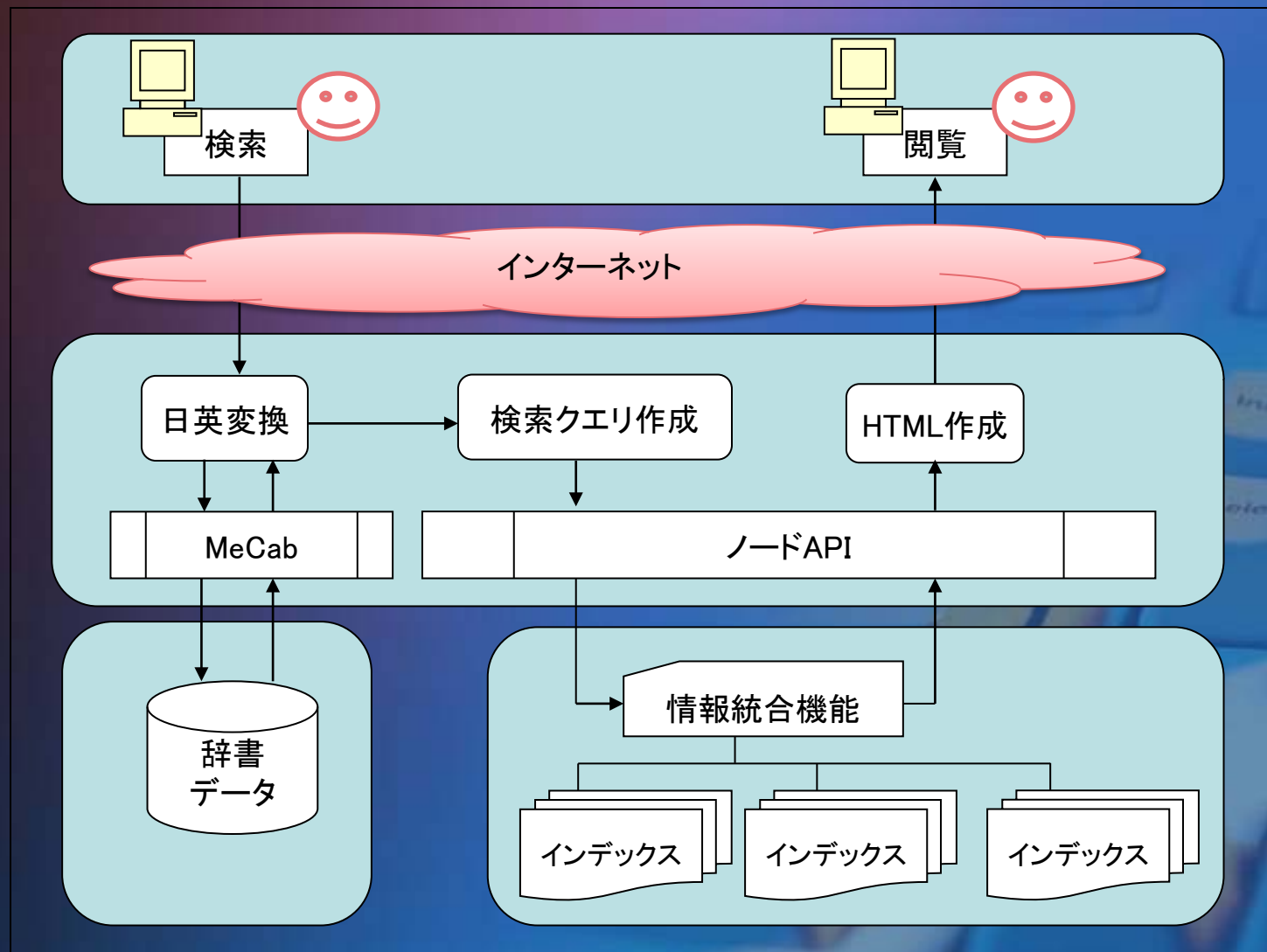
様々なアルゴリズムによるプログラムが開発された

生命科学データベース横断検索サービスの特徴

- 国内中心に集めた200以上の分子DBや文献をキーワードで検索できるGoogleライクな検索です
- 分子だけでなくPNEや特許などの文献も同時に検索できます
- 検索エンジンとしてhyperestraierを利用しています
- 日本語文書のインデックスはnグラム法で作製しています
- 日英翻訳が可能でライフサイエンス辞書(LSD)を使っています
- DBをカテゴリーごとに分けているので興味のある分野の絞り込みが簡単です

1月から著作権の一部が改正され検索エンジンが合法になりました

生命科学データベース横断検索エンジンの概略



様々な検索:あの人検索 スパイシー

あのひと検索 スパイシー

SPYSEE

大久保公策

人物検索

アドバンス検索

おすすめ

松原謙二

高木利久

坊農秀雅

川本祥子

五條堀孝

ユーザ登録 | ログイン

アスビー!

日本代表チーム

応援キャンペーン

大久保公策



hirogenome-sci.jp

画像について

この画像を表示しない







もっと

情報の戦略的な活用

特長的な競争優位のための、情報の戦略的な活用をIBMが支援。
 IBM.com

遺伝子解析受託(次世代他)

Illumina, Affymetrix 公認プロバイダー GLP認証(TK&試験) 施設 GPバイオ
 www.gpbio.jp

DNAマイクロレイ受託

極微量対応 解析ソフトも無償提供 トータルサポートのフィルジェン(株)
 www.filgen.jp

Ads by Google



Ads by Google

ウェブ上からみた人物像

プロフィールをもっと見る・変更する

プロフィール氏名大久保公策教授所属国立遺伝学研究所 / 総合研究大学院大学研究室生命情報・DOBJ研究センター遺伝子発現解析研究室研究テーマゲノムワイドな測定からの知識発見 公的データおよび知識データの動員を行なってヒト遺伝子群の構造化と知識発見同機関の研究室紹介小原 雄治所長「遺伝研はライフサイエンスの原点」荒木 弘之教授「研究は情熱と覚悟でできている」一色 孝子助教教授「しなやかな強さを秘めた研究者さん できた全ての中で一番好きだか」ら大久保 公策教授「ライフサイエンスの環境を整備する」角谷 徹に教...
<http://www.incu-be.com/myplace/academ...> より抜粋

DNA塩基配列解析は

最安・最速・高品質 700~850円 DNA/次世代塩基配列解析はこちら！

J-Bio21の遺伝子解析受託

審核されたノウハウを元に、研究、遺伝子検査のお手伝いを致します。

Ads by Google

相関図

もっと大きく表示



こんな人と関係があります

もっと見る

田野保雄

本巻1号の「理事会から」で田野保雄理事長が日本眼科学会戦略企画会議の発会について...

五條堀孝

日本の遺伝学者。

高木利久

高木利久[タカギトシヒサ] 1976年東京大学工学部計数工学科卒業。九州大学を...

小原雄治

氏名小原 雄治(こはら ゆうじ)所属国立遺伝学研究所 / 総合研究大学院大学研究...

タグ

学者・研究者

国立遺伝学研究所

バイオインフォマティクス

タグを編集するには [ログイン](#) してください。 [ユーザ登録はこちら](#)
 不適切なページとして報告する

検索結果に付加情報をつける： レコメンデーション 推薦システム



改訂第2版
バイオデータベースとウェブツールの手とり足とり活用法
遺伝子の配列・機能解析、タンパク質解析、プロテオミクス、文献検索、検索エンジン...etc.真に役立つサイトを使い倒す！
著者：中村 保一、石川 淳、鎌倉 敦、平川 美夏、坊間 秀雅
※本書は、Amazon.co.jpで販売、発送します。ギフト包装を利用できます。

その他のイメージを見る

自分のイメージを掲載する

出版社、著者の方へ「なが見！検索」で書籍を紹介させよう

改訂第2版 バイオデータベースとウェブツールの手とり足とり活用法 — 遺伝子の配列・機能解析、タンパク質解析、プロテオミクス、文献検索、検索エンジン...etc.真に役立つサイトを使い倒す！ (大型本)

中村 保一 (編集), 石川 淳 (編集), 鎌倉 敦 (編集), 平川 美夏 (編集), 坊間 秀雅 (編集)

またカスタマーレビューはありません。最初のレビューを書く

価格: **¥ 4,410** 国内配送料無料(一部例外あり) [詳細](#)

ポイント: 44pt (1%) [詳細はこちら](#)

在庫あり。 [在庫状況](#)について
この商品は、Amazon.co.jp が販売、発送します。ギフト包装を利用できます。

2009/1/24 土曜日にお届けします！ 今から20時間と25分以内にレジに進み、「お急ぎ便」オプション(有料)を選択して注文を確定されたご注文が対象です。詳しくは[こちら](#)
一部サイズが大きい商品の場合、上記の日付が適用されない場合があります。配送オプションを選択する画面、もしくは注文確定するときに必ず配送予定日をご確認ください。

新品1点 ¥ 4,410より

[予約注文・限定版／初回版・特典に関する注意](#)

[Click here to see in English.](#)

数量: 1

または
1-Clickで注文する場合は、[サインイン](#)をしてください。

こちらからも買えますよ

この商品を安く買いたい！
[中古商品を予約する](#)

この商品をお持ちですか？
[マーケットプレイスに出品する](#)

[ほしい物リストに追加する](#)

[お友だちに知らせる](#)

これを買ったひとはこっちも買ってます

この商品をあわせて買う



+



合計価格: **¥ 8,820**
ポイント合計: 88pt (1%)

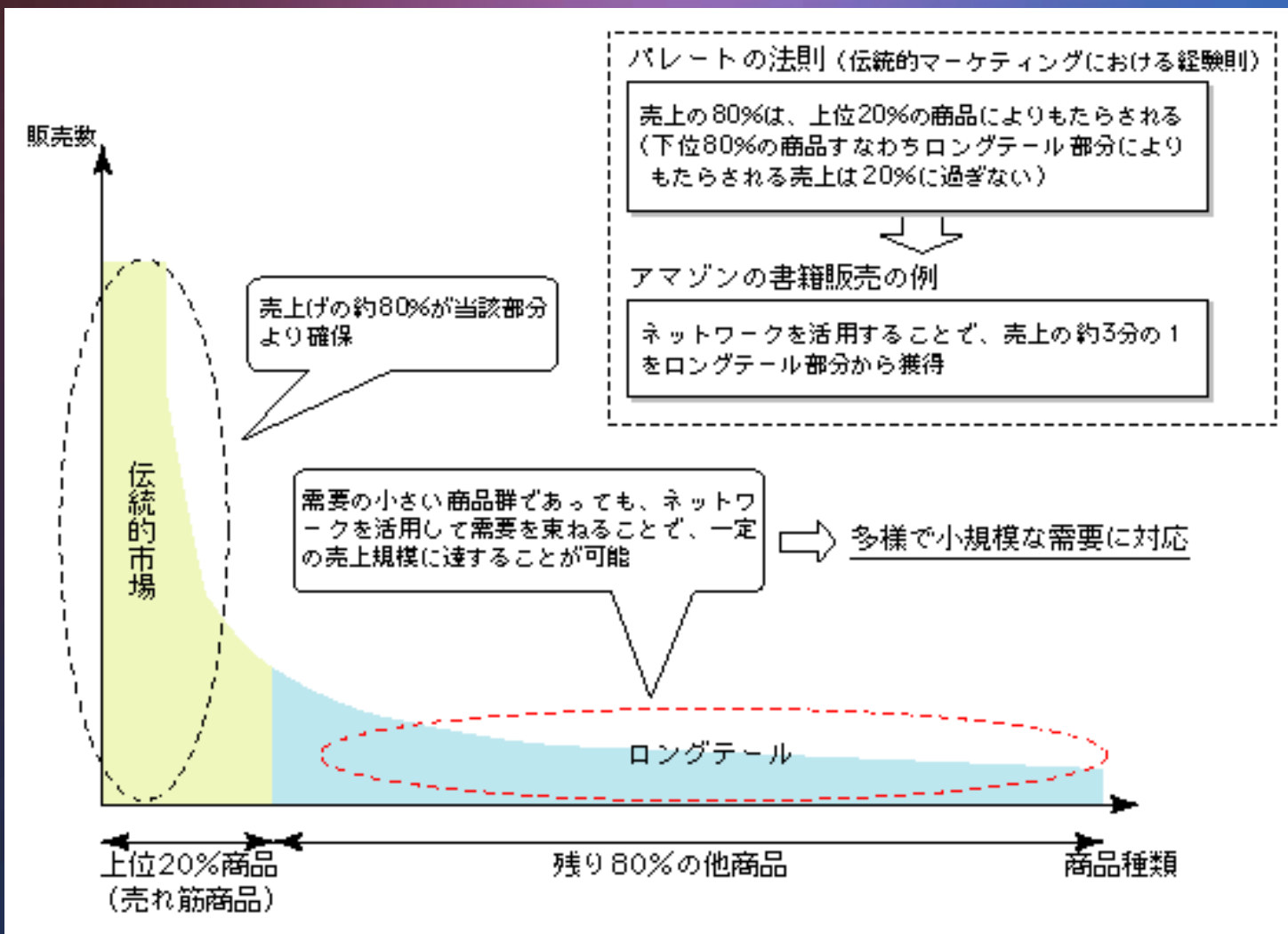
[両方ともカートに入れる](#)

この商品を買った人はこんな商品も買っています

ページ: 1 / 20



ロングテール現象



ウェブ技術を組み合わせた情報検索の例： マッシュアップ



- ◎ インフルエンザなど感染症の動向を探る
- ◎ Googleの検索キーワードの傾向を見ることが有効
- ◎ Twitterリアルタイム検索にGoogleが対応

文書に意味をもたせる

検索者が欲しい文書をより良く探せるように

- Xmlなどのファイル形式
- RDFなどのメタデータ
- オントロジー・推論

セマンティック ウェブ

WWWの生みの親が提唱
Tim Berners Lee

