# 生命科学データベース横断検索の現状

○三橋信孝[1]、杉崎太一朗[2]、 牧口大旭[2]、今西麗子[2]、奥村利幸[2]、川本祥子[3]

1.独立行政法人 科学技術振興機構 バイオサイエンスデータベースセンター(NBDC)、2.三井情報株式会社、
3.大学共同利用法人 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 (DBCLS)

## 生命科学データベース横断検索とは？

**散在する生命科学分野の分子データをはじめとするファクトデータ、特許、文献をキーワードで一括検索できるサービス**

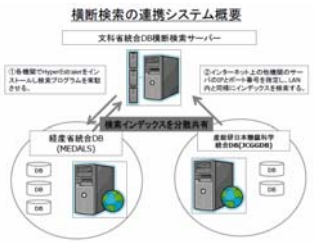
### サービスの特徴

- データベースカタログに収録されたものを中心に約270のデータベースが検索対象
- 各データベースサイトを訪問する手間を省ける
- データベース分類を検索結果の絞り込みに利用でき、知りたい分野の検索結果のみを抽出可能
- キーワード翻訳機能(ライフサイエンス辞書(京都大学)利用)により、日本語キーワードでも英語ページがヒット(英→日も)
- 日本語コンテンツの充実
  - 蛋白質核酸酵素のバックナンバー(全文(1985年～2008年)、抄録(1985年～2010年))
  - 学会要旨(日本生物物理学会)
  - 日本国特許(公開広報、特許実用新案広報の全文(2004年～))
  - 文科省や経産省「統合データベースプロジェクト」の成果物など

http://biosciencedbc.jp/dbsearch

### システムの特徴

全文検索エンジンにフリーソフトウェアHyperEstraierを利用して分散型のシステムを構築。各研究機関が各々更新するDBをHyperEstraierのP2P連携機能を用いてインターネット上で仮想的に共有し、横断検索を実現。研究機関の追加も容易。




### これまでの経緯

2008年6月： 文科省「統合データベースプロジェクト」の一環として、DBCLSがサービス開始。
2008年8月： 産総研糖鎖医工学研究センターの日本糖鎖科学データベース(JCGGDB)をP2P連携で追加。
2010年5月： 経産省「統合データベースプロジェクト」のMEDALS横断検索とデータベース検索の相互利用開始。
2011年4月： 運用業務をNBDCが担当。DBCLSはRDF化や高機能化等の研究開発を中心に。
2011年9月： 英語版公開

## 現在・今後の取り組み

**Googleとは違う、生命科学向けのきめ細やかな検索サービスを目指して**

### データ網羅性向上のため

#### Linked Dataの取り込み

- 今後、DBCLSが実施中の「基盤技術開発」でRDF化されていくデータ
- 公開されているRDFデータ(Linking Open Drug Dataなど)

### 連携機関の拡大

- 経済産業省に加え、農林水産省、厚生労働省のデータベースも

MEDALS (経済産業省・産業技術総合研究所) / JCGGDB / AgriTOGO (農林水産省・農業生物資源研究所) / NIBIO (厚生労働省・医薬基盤研究所)

- 「統合化推進プロジェクト」10研究課題との連携
- NBDCカタログデータベースに追加されていくデータベースの取り込み

#### 手軽に横断検索サーバを構築

- VPS (Virtual Private Server) やクラウド利用の試み
- サーバ構築・運用のマニュアル化

### データ適合性向上のため (生命科学分野に限定)

### データ最新性向上のため

#### インデックス更新の効率化

- 対象サイト更新有無の自動検知
- 対象サイトの更新頻度に応じたインデックス更新作業

### データのフィルタリング

特許データ(公開広報、特許・実用新案広報)をIPC分類を用いて、以下の生命科学関連カテゴリに限定

| | |
|---|---|
| A61K | 医薬用製剤 |
| A61P | 化合物または医薬組成物の治療活性 |
| C07C | 非環式化合物または炭素環式化合物 |
| C07D | 複素環式化合物 |
| C07H | 糖類;その誘導体;ヌクレオシド;ヌクレオチド;核酸 |
| C07K | ペプチド |
| C12 | 生化学;ビール;酒精;ぶどう酢;酢;微生物学;酵素学;突然変異または遺伝子工学 |
| G01 | 測定;試験 |
| G07N | 材料の化学的または物理的性質の決定による材料の調査または分析 |

他、Wikipediaでも生命科学関連エントリの抽出に挑戦中

トーゴーの日シンポジウム 2011 ライフサイエンス分野のデータベース統合の“カタチ”を探る
会期：2011.10.5 於：日本科学未来館 みらい CAN ホール

Licensed under a Creative Commons 表示2.1日本 license
©2011 三橋信孝 バイオサイエンスデータベースセンター