統合データベースプロジェクト　シンポジウム2009

日本語による質問で
世界の生命科学情報を検索するために
-デジタル対訳辞書の作成 -

大学共同利用機関法人　情報･システム研究機構
(1)ライフサイエンス統合データベースセンター
(2)DDBJ/生命情報研究センター
(3)日立ソフトウェアエンジニアリング株式会社

〇大久保公策1)2),栗原英輔3),渡邊康司 3), 森山丈士3),
小笠原理2), 川本祥子1)

デジタル対訳辞書が何故必要か

- ２１世紀の科学ではデジタル革命による科学の推進がテーマです
- 自由に利用できる科学情報がインターネットを通じて世界を満たします

デジタル対訳辞書が何故必要か

- 公的科学データの共有は世界全体で実現するべき課題 (04年OECD声明)

- しかし日本語を母国語とする我が国では、豊富なデジタルデータを利用することに言葉の障壁が存在します

デジタル対訳が何故必要か

- ライフサイエンス統合DBプロジェクトでは

  1）日本語の質問で世界の科学情報（答え）に到達可能にする

  2）日本語で書かれた公的科学文書もデジタル化し検索可能にする

以上の二つの努力を行っています。

デジタル対訳辞書の特徴 その1

- 意味に踏み込まない
  - 対訳辞書は意味の表現を含みません
  - また、特定の外国語に対してどのような日本語を使うべきかを決めたものではありません
  - どのような日本語が「使われているか」を信頼性のおける手法で調べた「データ」です

デジタル対訳辞書の特徴 その2

- 分野別でない
  - 分野の垣根を越えたセレンディピティを期待される検索エンジンのために対訳辞書も細かい専門領域を超えた辞書が望まれます
  - 領域別に編纂されたものを共通のデータ形式に再編します

デジタル対訳辞書の特徴 その3

- 利用制約が少ない
  - 紙の辞書は、意味を提供するなどして人に直接「読まれる」ことが主ですが、意味を含まないデジタル辞書は、検索その他のサービスを提供する際に「コンピュータのプログラム」で利用する素材です。
  - 素材の加工や組み合わせが制約無く自由に出来ることによって真価を発揮し、より多くの価値が生み出されます。

デジタル対訳辞書の特徴 その4

- 情報源
  - 各分野の専門家や学会が過去に出版した学術用語集、辞書、図鑑などから、専門用語とその対訳部分を取り出します
  - 出版物は著作物として著作権法により守られており、DBC辞書は著作権を侵害しない範囲で作成配布を行います

（→著作権問題を専門とする法律家のコメント参照）

対訳を本から写し取ることは著作権の侵害にあたりますか？

- 意味解説や例文や用法の表現など、自由度の高い表現には著作権があり、許諾を得ずにそのまま「引用」すると問題となる可能性があります。しかしながら、英語の単語は万民の共有財産であって著作権がなく、また、これを日本語にした専門用語の対訳には、表現の自由度は極めて少なく、これもまた共有財産となるものです。訳語とは、本来、訳者の個性を表現することが求められるのではなく、むしろ、すべての科学者が同じ用語を用いるという統一利用が期待されています。すなわち専門用語の正しい対訳には「創作性」を発揮する余地はないと考えられます。専門用語辞書を発行するにあたっては、その辞書に収録すべき用語の選定にあたって、編集著作物としての創作性が存在することはおおいに考えられます。しかし、この対訳辞書は、既存の辞書の編集方針に従うことなく、独自の視点で、広く多くの分野にまたがった専門用語を選択し、その対訳の選択にあたっても独自の編集方針に基づいて行っており、編集著作物としては独立のものです。

（著作権問題を専門とする法律家のコメント）

著作権以外の法的制約はありますか？

- 一般に、書籍の末尾に「無断転載を禁止します」等の記載がある場合には、その趣旨は、その出版物に存在している著作権に基づいて、無断転載など、著作権法が禁止している行為を行わないよう注意を喚起しているものと理解されます。デジタル化を禁止しているものも、「デジタル化権」という権利が存在していないと解釈されている現在の状況では、著作物の複製を禁止しているものと解釈できます。したがって、著作権の侵害とならない行為までが、このような書籍の末尾の記載によって禁止されるものではありません。また、書籍の末尾に記載されている文言は、一般に注意文言と解釈されており、契約を構成するものではありません。契約を構成するためには、契約の申込みと、それに対する承諾が必要ですが、書籍の末尾に記載されているものは、通常の契約書に両当事者が署名捺印する場合や、上記の簡易契約のうち、契約条件を提示された上でクリックボタンを押す場合などと異なり、申込みと承諾、という行為が存在していないため、契約とは考えることができないためです。

（著作権問題を専門とする法律家のコメント）

原典の著者や出版社への配慮

- 出版物を原著として編纂したデジタル辞書の流通が出版社の専門用語辞書編纂販売に与える影響には細心の配慮をする必要があると考えます
- 出版物の辞書は図表を含む用語解説や配列の方法などの創造的な表現によってそれぞれ個性を出していますが、ここで開発する辞書は用語解説を含まず、また配列においても多数の分野から独自の手法でその用語のみを再編集したものであり、その用途は全く異なっているため、競合するものではありません
- デジタル辞書の配布にあたっては　原典と著者を明記し、またその情報が常に表示されるなど、オリジナル編纂の努力と知力がクレジットとして、デジタルデータと共に流通する形で敬意を払う方法が望ましいと考えます

デジタル対訳生物学名について

- 和名対訳辞書作成方針
  - 生物種およびクラス名称についてラテン名、英語名に対応した和名を情報源と共に集める
  - 生物種の構造化（分類）は行わない
  - 情報源による食い違いや表記ゆれは正当性の検討は行わず情報源の併記に留める
  - 公式に決められた和名（標準和名）を探す。探し出せなかった場合は、それに代わる和名を適切な情報源から探す

デジタル対訳生物学名について

- 想定アプリケーション
  - 検索サービスの質問用語を自動翻訳する
  - 分子データに対する 材料情報を和訳する
  - 論文の生物名索引を和訳する
  - その他生物多様性情報を日本語で利用可能にする

生物名称対訳用語集データの検索画面
サケで検索

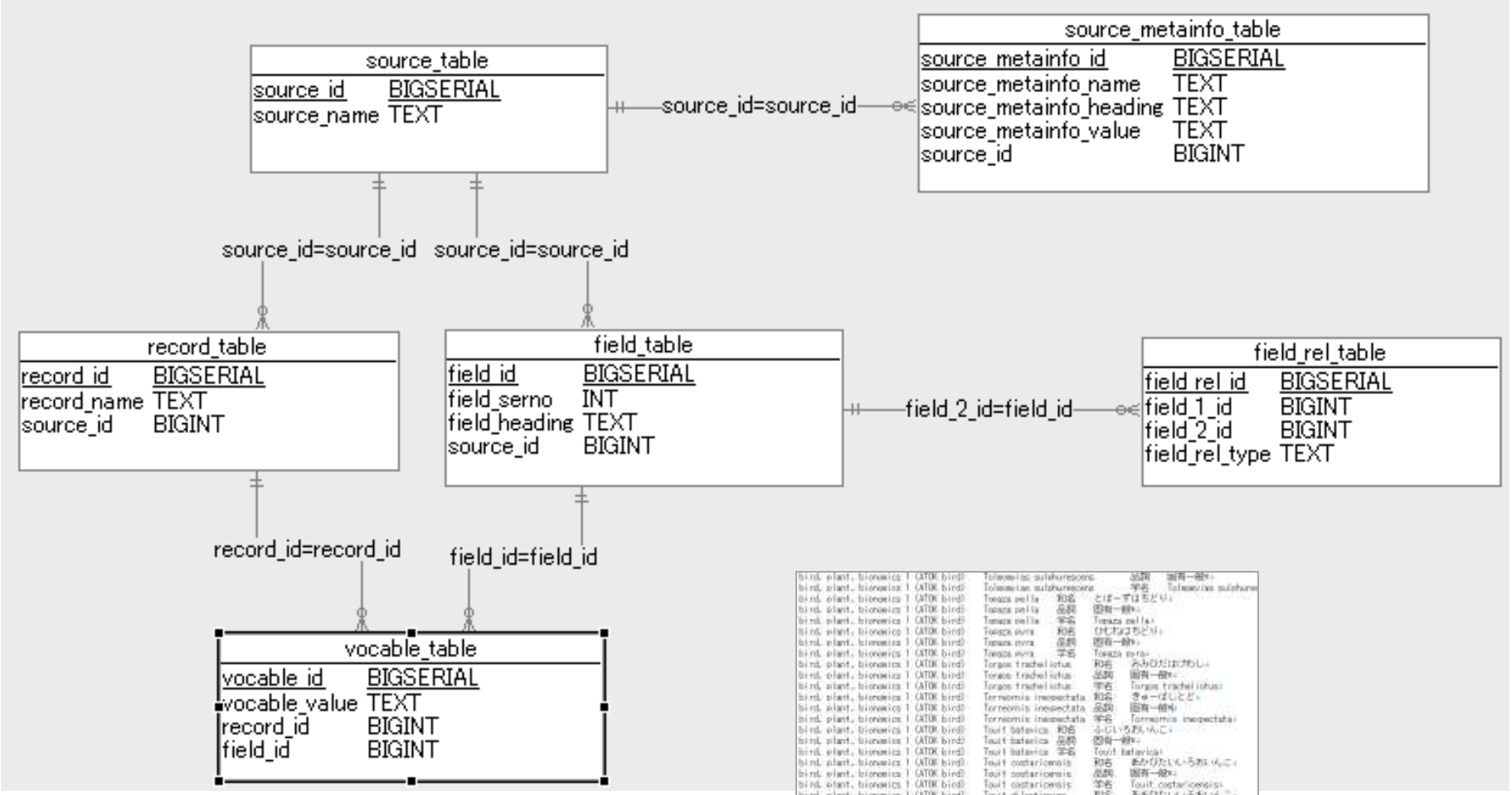
Species Dictionaries

サケ [search]

Hit count:115

```
dic/isbn_978-4-525-01501-5.bacteria.all.txt                                Lactobacillus sakei
 dic/isbn_978-4-525-01501-5.bacteria.all.txt                               Lactobacillus sakei subs
 dic/isbn_978-4-525-01501-5.bacteria.all.txt                               Lactobacillus sakei subs
 dic/株式会社北隆館_新訂原色昆虫大圖鑑第I巻_新訂初版_和名索引.txt                  アサケンモン        102
 dic/株式会社北隆館_新訂原色昆虫大圖鑑第III巻_新訂初版_和名索引.txt                        イトウサケオオドリバエ
 dic/株式会社北隆館_新訂原色昆虫大圖鑑第III巻_新訂初版_和名索引.txt                        クロサケオオドリバエ
 dic/sciterms_zoo91.species_names.txt   0185650 サケ目    Sakemoku        サケモク   Order Salmonifor
 dic/sciterms_zoo91.species_names.txt   0185710 サケスズキ目       Sakesuzukimoku  サケスズキモク
 dic/sciterms_zoo91.species_names.txt   0186850 サケビドリ亜目      Sakebidoriamoku サケビドリアモク
 dic/sciterms_zoo91.species_names.txt   0187050 サケイ亜目 Sakeiamoku      サケイアモク        Suborder
 dic/PLANTS_ORCHID_Plam_leaf_spd.dic.txt         Androsace L.     アンドロサケ
 dic/BIRDC.dic.txt       Anhima cornuta  ツノサケビドリ
 dic/atok_dics.dic.txt   Anhima cornuta  ツノサケビドリ
 dic/atok_dics.dic.txt   Black-bellied Sandgrouse         クロハラサケイ
 dic/atok_dics.dic.txt   Black-faced Sandgrouse  クロガオサケイ
 dic/niaes.dic.txt       Blumea laciniata (Roxb.) DC.     サケバコウゾリナ
 dic/atok_dics.dic.txt   Burchell's Sandgrouse   シロボシサケイ
 dic/BIRDC.dic.txt       Chauna chavaria クロエリサケビドリ
 dic/atok_dics.dic.txt   Chauna chavaria クロエリサケビドリ
 dic/BIRDC.dic.txt       Chauna torquata ガンムリサケビドリ
 dic/atok_dics.dic.txt   Chauna torquata ガンムリサケビドリ
 dic/atok_dics.dic.txt   Chestnut-bellied Sandgrouse      チャバラサケイ
 dic/wikipedia.script.txt        Circinella       カラクサケケカビ属
 dic/atok_dics.dic.txt   Crowned Sandgrouse       ササフサケイ
 dic/DJI.dic.txt         Diamesa plumicornis      フサケヤマユスリカ
 dic/DJI.dic.txt         Dirades theclatus        アサケマルバネフタオ
 dic/atok_dics.dic.txt   Double-banded Sandgrouse         フタオビサケイ
 dic/atok_dics.dic.txt   Four-banded Sandgrouse   ヨツオビサケイ
 dic/PLANTS_ORCHID_Plam_leaf_spd.dic.txt         Fumaria officinalis      カラクサケマン
 dic/DJI.dic.txt         Glipostenoda excisa      サケオヒメハナノミ
 dic/DJI.dic.txt         Glipostenoda pseudexcisa         ニセサケオヒメハナノミ
 dic/atok_dics.dic.txt   Horned Screamer ツノサケビドリ
 dic/PLANTS_ORCHID_Plam_leaf_spd.dic.txt         Jatropha multifida       サケバヤトロファ
 dic/PLANTS_ORCHID_Plam_leaf_spd.dic.txt         Jessenia oligocarpa Griseb. & H.Wendl.   オニサケヤシ
 dic/PLANTS_ORCHID_Plam_leaf_spd.dic.txt         Jubaea chilensis         チリサケヤシ
 dic/PLANTS_ORCHID_Plam_leaf_spd.dic.txt         Jubaea chilensis (Molina) Baill.         チリサケヤシ
 dic/isbn_978-4-525-01501-5.bacteria-katakana.txt        Lactobacillus sakei      262      Katagiri
 dic/isbn_978-4-525-01501-5.bacteria-katakana.txt        Lactobacillus sakei subsp. sakei
```

現在の辞書の構造


source_table
source_id BIGSERIAL
source_name TEXT
source_id=source_id
source_metainfo_table
source_metainfo_id BIGSERIAL
source_metainfo_name TEXT
source_metainfo_heading TEXT
source_metainfo_value TEXT
source_id BIGINT
source_id=source_id
source_id=source_id
record_table
record_id BIGSERIAL
record_name TEXT
source_id BIGINT
field_table
field_id BIGSERIAL
field_serno INT
field_heading TEXT
source_id BIGINT
field_2_id=field_id
field_rel_table
field_rel_id BIGSERIAL
field_1_id BIGINT
field_2_id BIGINT
field_rel_type TEXT
record_id=record_id
field_id=field_id
vocable_table
vocable_id BIGSERIAL
vocable_value TEXT
record_id BIGINT
field_id BIGINT

Ex.

情報源のリスト（抜粋）

- 書籍
  - 南山堂「微生物学用語集 英和・和英」(isbn:978-4-525-01501-5)
    - ソース/生物名/isbn_978-4-525-01501-5.bacteria.txt
    - ソース/生物名/isbn_978-4-525-01501-5.fungi.txt
    - ソース/生物名/isbn_978-4-525-01501-5.virus.txt
    - ソース/生物名/isbn_978-4-525-01501-5.bacteria-katakana.txt
    - ソース/生物名/isbn_978-4-525-01501-5.bacteria.all.txt
    - ソース/生物名/微生物学用語集の納品物
    - 読み取り外注/微生物名/微生物学用語集の原稿
  - 大学書林「世界鳥類和名辞典」山階 芳麿 (著) (ISBN-13: 978-4475001151)
    - ソース/生物名/鳥類学名データ.鳥類本文.txt
    - ソース/生物名/鳥類学名データ.鳥類分布.txt
    - ソース/生物名/BIRDC.dic.txt **(別の情報源の混入あり)**
    - 書籍のコピー/isbn_4475001153.1.pdf
    - 書籍のコピー/isbn_4475001153.2.pdf
  - 原書房「世界鳥類名検索辞典 和名篇」白井 祥平 (編著) (ISBN-13: 978-456202350
    - ソース/生物名/BIRDC.dic.txt **(別の情報源の混入あり)**
  - 岩波書店「岩波 生物学辞典 第 3 版」(isbn:4-00-080018-3)
    - ソース/生物名/isbn_4_00_080018_3_virus.txt
    - ソース/生物名/isbn_4_00_080018_3_bacteria.txt
    - ソース/生物名/isbn_4_00_080018_3_mycota.txt
    - ソース/生物名/isbn_4_00_080018_3.pdf
  - 平凡社「世界哺乳類和名事典」(isbn: 4-256-10711-8)
    - ソース/生物名/mammal.isbn_4_256_10711_8.txt
    - ソース/生物名/哺乳類辞典cap.txt
    - ソース/生物名/哺乳類辞典cap.txt(誤字レベルの補正など)
    - ソース/生物名/哺乳類辞典cap.txt(原稿中の誤字訂正や構文解析など)
    - ソース/生物名/平凡社.世界哺乳類和名辞典.isbn.4-256-10711-8.pdf
    - ソース/生物名/平凡社.世界哺乳類和名辞典.isbn.4-256-10711-8.書誌情報.pdf
  - 丸善「解剖学用語」**(版は不明)**
    - 読み取り外注/解剖学用語/解剖学用語の納品物
  - 北隆館「新訂 原色昆虫大圖鑑 第 I 巻 (蝶・蛾 篇)」(isbn: 978-4-8326-0825-2)
    - ソース/生物名/和名索引を書き写した納品物:「新訂 原色昆虫大圖鑑 第 I 巻」
    - ソース/生物名/学名索引を書き写した納品物:「新訂 原色昆虫大圖鑑 第 I 巻」
  - 北隆館「新訂 原色昆虫大圖鑑 第 II 巻 (甲虫 篇)」(isbn: 978-4-8326-0826-9)

- 厚生労働省検疫所「届出対象動物 種名、学名リスト」(http://iaa.keneki.jp/newpage5.html)
  - ソース/生物名/keneki_jp.dic.txt
- 福井県安全環境部自然保護課「福井県みどりのデータバンク」(http://www.erc.pref.fukui.jp/gbank/)
  - ソース/生物名/fukui_all.dic.txt
  - ソース(全カラム)／生物名／fukui_all.dic.txt／amph.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／bird.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／fr_shell.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／fukui_rdb1.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／fukui_rdb2.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／insect1.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／insect2.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／insect3.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／insect4.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／insect5.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／mammal.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／plant.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／rdb.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／rdbplant.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／reptile.csv
  - ソース(全カラム)／生物名／fukui_all.dic.txt／shell2.csv
- FishBase Consortium「FishBase」「Common names in 日本語 (Japanese)」(http://www.fisht
  - ソース/生物名/fish_list.dic.txt
- 国立情報学研究所「オンライン学術用語集」(http://sciterm.nii.ac.jp/cgi-bin/reference.cgi)「動物学編
  - ソース/生物名/sciterms_zoo91.species_names.txt
  - ソース/生物名/zoo91.sgml.tar.gz
  - ソース/生物名/bot92.sgml.tar.gz
  - ソース/生物名/gen93.sgml.tar.gz
- 山形大学理学部生物学科生物多様性大講座動物生態学研究室「学名辞書」(http://sci.kj.yamagata-u.a
  - ソース/生物名/atok_dics.dic.txt **(別の情報源の混入あり)**
- 環境省自然環境局生物多様性センター「生物多様性情報システム(J-IBIS)」「第 4 回自然環境保全基礎調査
  - ソース/生物名/biodic_go_jp.dic.txt
- 千葉大学理学部生物学科多様性生物学講座系統学研究分野系統分類学研究室「BG Plants 和名 - 学名イ
  - ソース/生物名/atok_dics.dic.txt **(別の情報源の混入あり)**
- 財団法人実験動物中央研究所微生物検査部「ICLAS」「微生物カテゴリーわけ」(http://www.iclasmonic.
  - ソース/生物名/ICLAS.dic.txt
- DDBJ から DDBJ release 64 登録塩基数の多い生物上位 100
  - ソース/生物名/ddbj.dic.txt
- The International Seed Federation (ISF) から植物病害関連生物リスト