

# 名前はなぜ データベースの 問題となるのか

筑波大学 伊藤 希



[www.nomencurator.org](http://www.nomencurator.org)



# 目次

1. 名前が難しい?
2. 名前とは何か
3. 学名・和名・一般名
4. データベースにおける名前
5. 生物名データベースの現状



# 目次

1. 名前が難しい?
2. 名前とは何か
3. 学名・和名・一般名
4. データベースにおける名前
5. 生物名データベースの現状



[www.nomencurator.org](http://www.nomencurator.org)



# 名前が難しい？

名前を使わない日はなく、使えて当然  
名前を適切に使う訓練を受けてきている  
意識していないが故に、何が行なわれているか  
明文化されていない

名前の大規模処理にコンピュータは不可欠  
コンピュータは暗黙の処理が苦手

名前のモデルが必要



# 目次

1. 名前が難しい?
2. 名前とは何か
3. 学名・和名・一般名
4. データベースにおける名前
5. 生物名データベースの現状



名前とは何か

名前とは名前の名前である

...そりゃそうだけど



[www.nomencurator.org](http://www.nomencurator.org)



# 名前とは何か

名前に求められること... 区別

ものが複数存在するから名前が必要

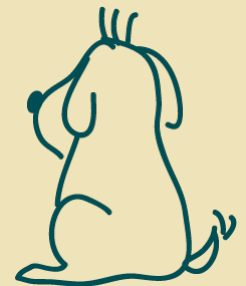
同じものは同じ名前、違うものは違う名前

同じであるとはどういうことか

表現の同一性?

指示対象の同一性?

指示対象が単一であること?



# 表現の同一性

## 同姓同名、異物同名の存在

文字列表現が同じでも名前として異なる場合

文字列として異なるなら別の名前か？

単純なスペルミス...学名ではしばしば見られる

colour と color は別名か？

言語が違くと異なる名前なのか？

命名規約、名前空間



[www.nomencurator.org](http://www.nomencurator.org)





# 指示対象の同一性

## 固有名の場合

指示対象は個物であり、同一性の判断は容易  
指示対象が単一であるかどうかで決まる

## 集合名の場合

指示対象が集合の場合、個物同様に扱える  
指示対象が概念である場合同一性の判断は必ずしも明白ではない



# 単一指示対象

## 明けの明星・宵の明星

名前が指示している対象は単一の惑星  
指示対象が同一なら入れ替えてよい？

「明けの明星は宵の明星と同じ惑星」

「明けの明星は明けの明星と同じ惑星」

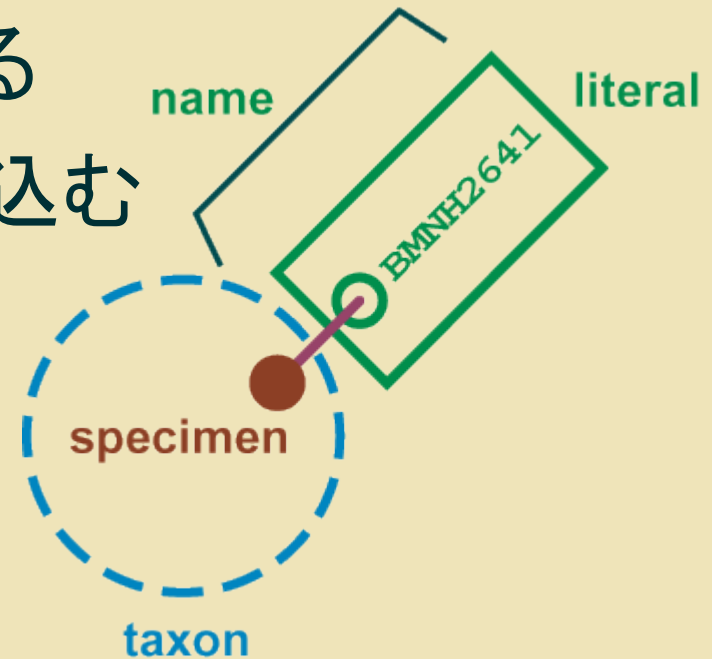
入れ替えて何かが失われるなら、同一の名ではない

対象の単一性以上の何かを担う場合がある

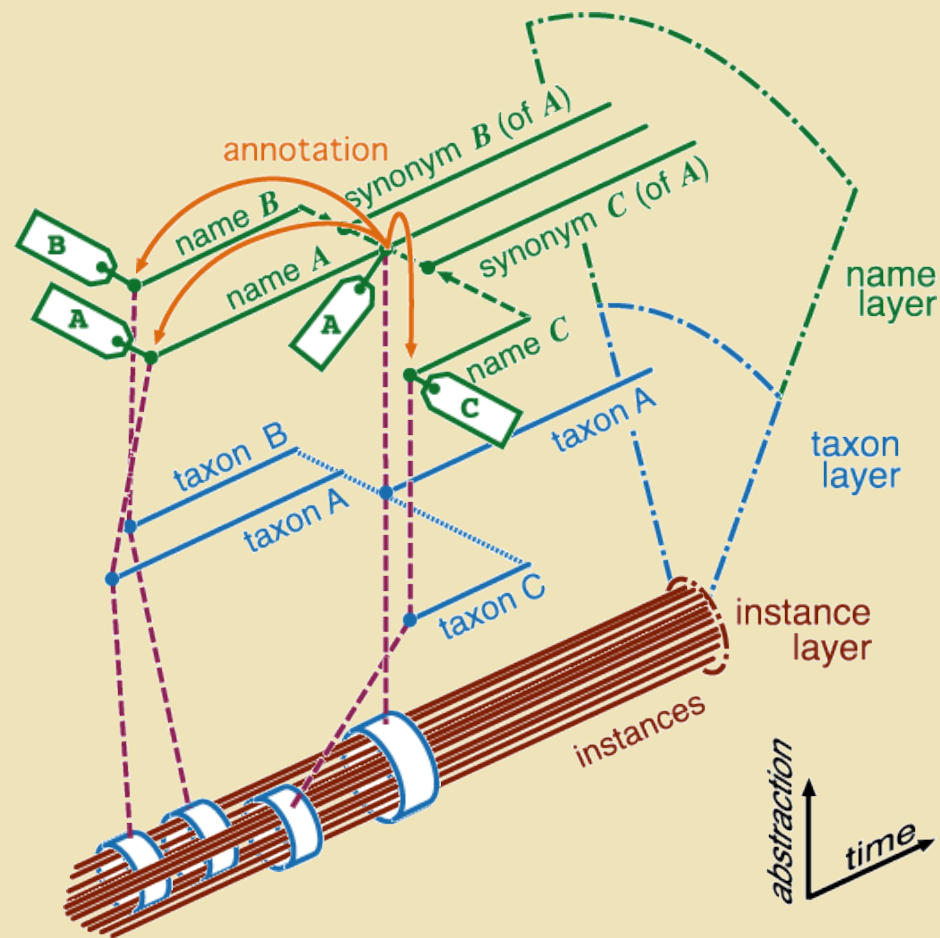


# 名前とは何か

名前とは表現そのものではない  
もちろん対象そのものでもない  
表現と対象の対で名前となる  
名前を扱うには対象に踏み込む  
必要がある  
時間的に変化し得る



# 分類学における名前の変遷



# 目次

1. 名前が難しい?
2. 名前とは何か
- 3. 学名・和名・一般名**
4. データベースにおける名前
5. 生物名データベースの現状



# 学名・和名・一般名

名前に求められること... 区別

通常は同一言語内での区別が問題

学名、和名、一般名は互いに異言語

それぞれが異なる有効範囲、異なる体系を  
有し、異なる対象を指示する可能性がある

生物名を扱うデータベースでは全てをカバー  
する必要がある



[www.nomencurator.org](http://www.nomencurator.org)



# 学名

命名規約が存在...規約ごとのスコープ

命名規約を超えたシノニム・ホモニムはない

属以上の分類群については単名法

種以下の分類群については二ないし三名法

種の記載に変化がなくとも、分類体系に依存して学名が変化し得る

指示対象だけでなく、分類体系も名前に含まれている...規約のスコープの含意でもある



[www.nomencurator.org](http://www.nomencurator.org)



# 和名

標準和名が存在

学会の定めたものとそうでないものがある

命名規約は存在しない

概ね種に対応するレベルで標準和名が与えられている点が特徴的

学名に比べ安定

差別用語排除などの例外あり



[www.nomencurator.org](http://www.nomencurator.org)





# 一般名

標準和名の如き標準名はむしろ稀

和名の様に種レベルに対応するのも稀

命名規約は存在しない

言語により当然異なり、方言や業界用語も  
存在



[www.nomencurator.org](http://www.nomencurator.org)



# 目次

1. 名前が難しい?
2. 名前とは何か
3. 学名・和名・一般名
4. データベースにおける名前
5. 生物名データベースの現状



[www.nomencurator.org](http://www.nomencurator.org)



# データベースにおける名前

名前に求められること... 区別

ものが複数存在するから名前が必要

現代のデータベースシステムは表の集まり

表が互いに参照して情報を表現している

たとえば、顧客リスト、商品リスト、発注リスト

個々の表には固有名がある

「主キー」で表の一行一行を区別する

主キーが行の「名前」になっている



# データベースにおける名前

名前が一意に決まるなら、単一テーブルの主キー管理で済む

異物同名や同物異名がある場合、別途対応表が必要になる

- 文字列の対応
- 表名前すなわち文字列と対象の対応表  
いずれかが必要

多言語対応には別の枠組み



# コンピュータでの多言語処理

ロカール (ロケール、locale) による  
言語、国、「方言」の組合せで言語等を指定

言語: ISO639-2, -3 ... ja

国: ISO3166 ... JP

「方言」: 特に規定なし

「ja\_JP.UTF-8」など

命名規約も言語の一種と言えなくもない

ICBN, ICZN, ...



[www.nomencurator.org](http://www.nomencurator.org)



# データベースにおける名前

名前 = (言語、文字列、対象) という三つ組  
で学名、和名とも名前として表現できる

単一言語対象なら「言語」指定は不要、文字列  
だけ扱うなら「対象」指定は不要

異物同名や同物異名がある場合、別途対応  
表が必要になる

「対象」は他テーブルの主キー、すなわち名  
前...問題の先送り?



# 目次

1. 名前が難しい?
2. 名前とは何か
3. 学名・和名・一般名
4. データベースにおける名前
5. 生物名データベースの現状



[www.nomencurator.org](http://www.nomencurator.org)



# 生物名データベース

## 二種類に大別

1. 生物名を表す文字列を集めたもの
  - a. 学名辞書
  - b. シノニムも含むシソーラス
  - c. 一般名も含む翻訳辞書
2. 生物の名前すなわち概念を考慮したもの



[www.nomencurator.org](http://www.nomencurator.org)





# 生物名DBの問題点

誰のためなのか、誰が維持管理するのか  
研究の進んでいる生物群は生物名データ  
ベースを維持管理する知見は蓄積整理さ  
れている

「生物名を表す文字列を集めたもの」作成可能  
分類群概念が整理されていれば「概念を考慮し  
たもの」も難しくはない (または、不要)

研究が進んでいない分類群ほどDBが必要



[www.nomencurator.org](http://www.nomencurator.org)



# 需要と供給

研究が進んでいない分類群について、需要と供給をどうマッチさせるか

供給できるもの: 文献OCR

Biodiversity Heritage Library

OCR 結果から生物名を抽出

名前の用例を集積

用例を対象のかわりにできれば...

OED方式: 用例を集めて辞書を作る



[www.nomencurator.org](http://www.nomencurator.org)



# 名前の用例で何ができそうか

名前に求められること... 区別

ものが複数存在するから名前が必要

単一文献内では名前は整合的に用いられているだろう (多少のミスはあるにしても)

名前の用例を適切に集積すれば、対象の異同位はわかるのでは

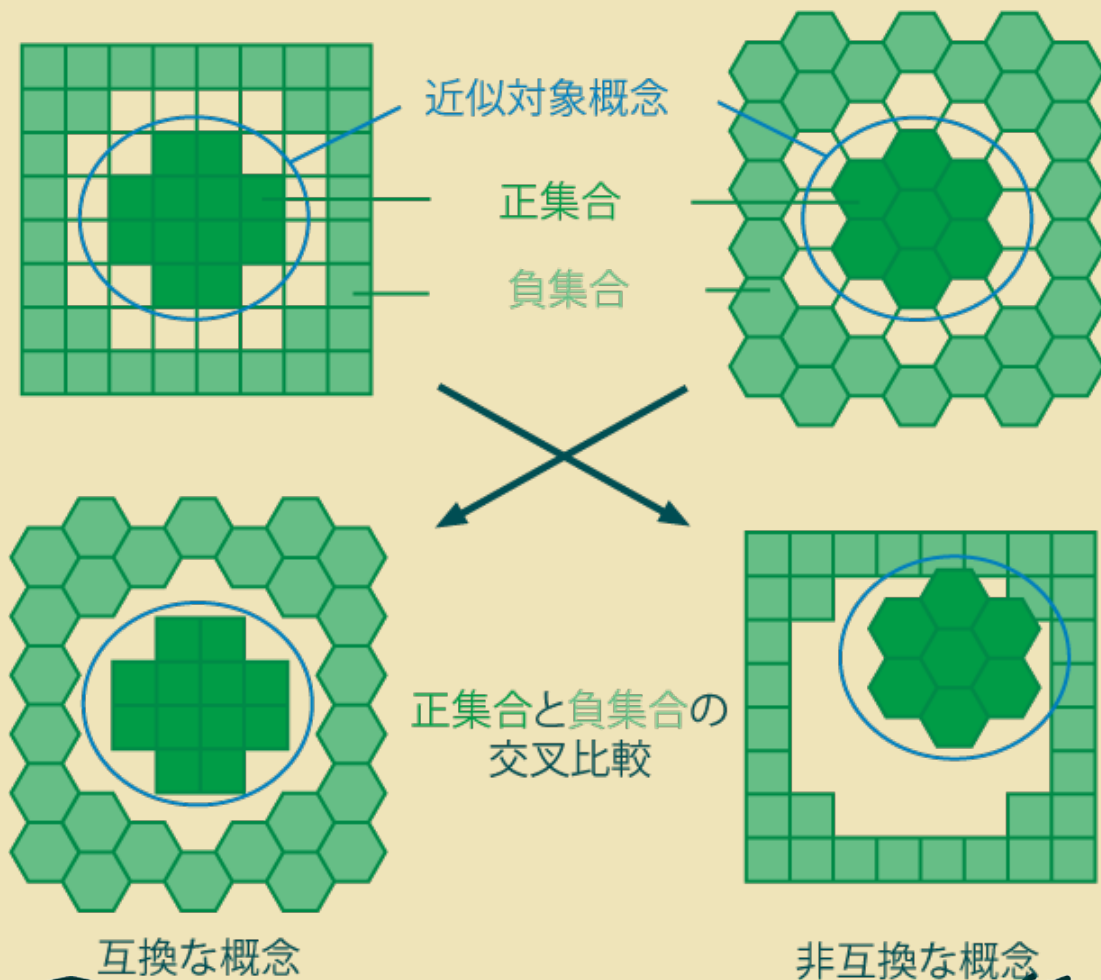
名前 = (言語、文字列、文献) を集積



[www.nomencurator.org](http://www.nomencurator.org)



# ラフ集合による概念比較



# 生物名DB関連アクティビティ

TDWG: TCS (Taxon Concept Schema)

生物名及び生物概念の情報交換標準

uBio.org: FindIT

学名抽出ツール

BHL: [www.biodiversitylibrary.org](http://www.biodiversitylibrary.org)

文献のOCRアーカイブ

GBIF: GNA (Global Names Architecture)

用例を含めた名前の集積 (開発進行中)



[www.nomencurator.org](http://www.nomencurator.org)

