



第23回統合DBミーティング

Integrated MTG in NIBIO

2013/8/30

本日の予定

- Sagace
 - アクセス解析
 - 進捗報告
 - ランキングの適用
 - クリックされないDBの絞り込みについて
 - 追加予定のDBの判断方法について
 - Sagaceに追加するDBについて
 - RDFa Liteとmicrodataについて
- 今後の予定

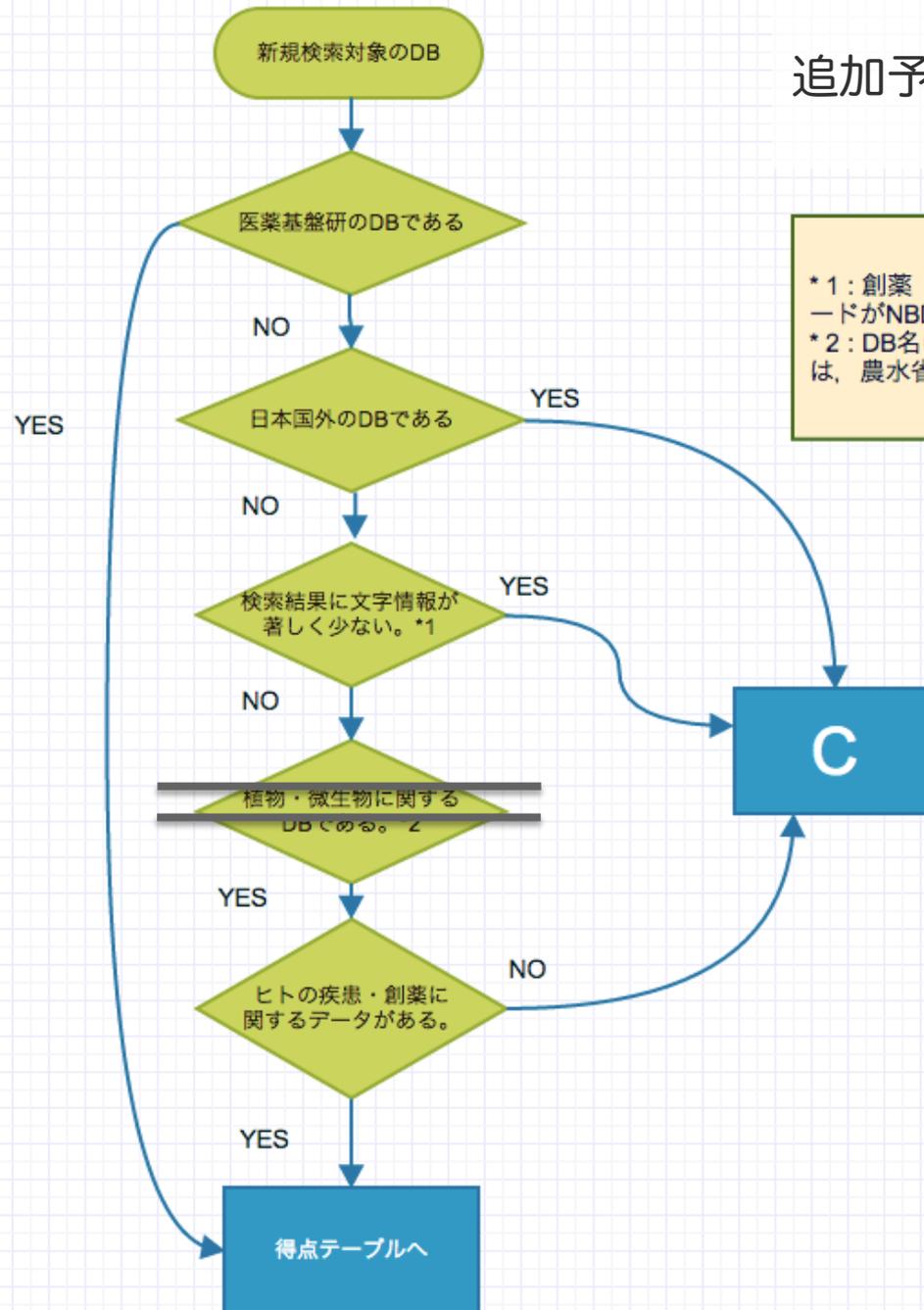
進捗報告

- ・ ランキング適用
 - ・ A:12000,B:8000
 - ・ 同じ日にクリックされたDBから重複を除き、実際にクリックされた最低順位のものだけを取り出す。
 - ・ 得点が低いもの(12400点以下) でクリックされたものは順位にかかわらず200点加点
 - ・ それ以外は, 10位までを100点, それ以下の順位を200点として加点していく
 - ・ 特許は8100点として固定
 - ・ 30000点以上にはならないようにする。

進捗報告 2

- ・ 検索されないDBの削除
 - 過去1回も検索されなかったDBを削除。
 - ・ 272 DB -> 122 DBへ
 - 0件時とサービスダウン時のページを作成

追加予定のDBの判断方法について



*1: 創薬 (化合物名も含む) ・疾患に関するキーワードがNBDC横断検索に表示されない。
*2: DB名に植物名・微生物名が含まれている。または、農水省系の機関のDBである。

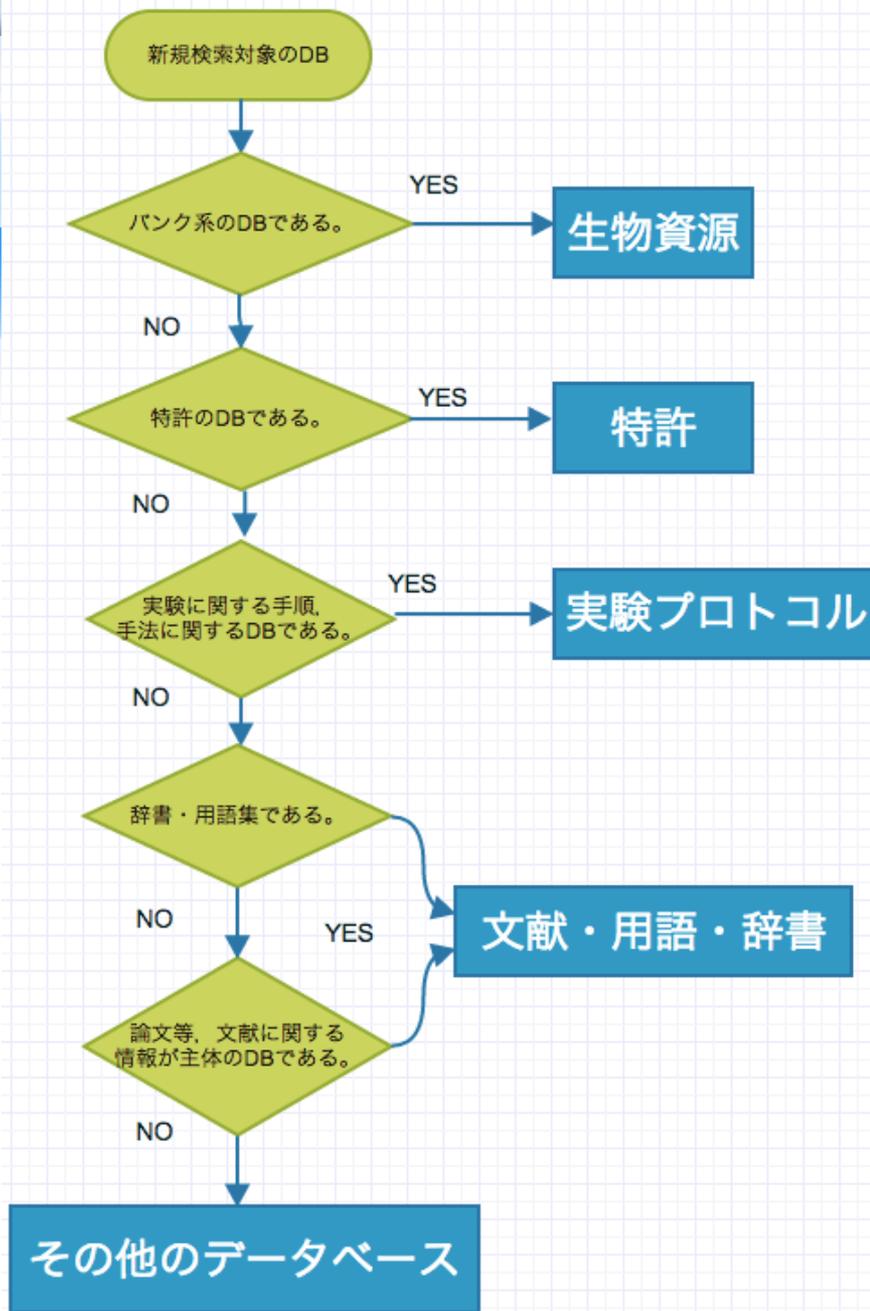
得点テーブル

創薬・疾患に直接関わるDBである。	2
文書系のDBである。	1
独自の実験・解析データを公開している。	1
研究者向けであり、 一般向けではない。	1
メタデータの取り組みに協力している。	2

合計得点が2点以上	A
合計得点が1点	B
合計得点が0点	C

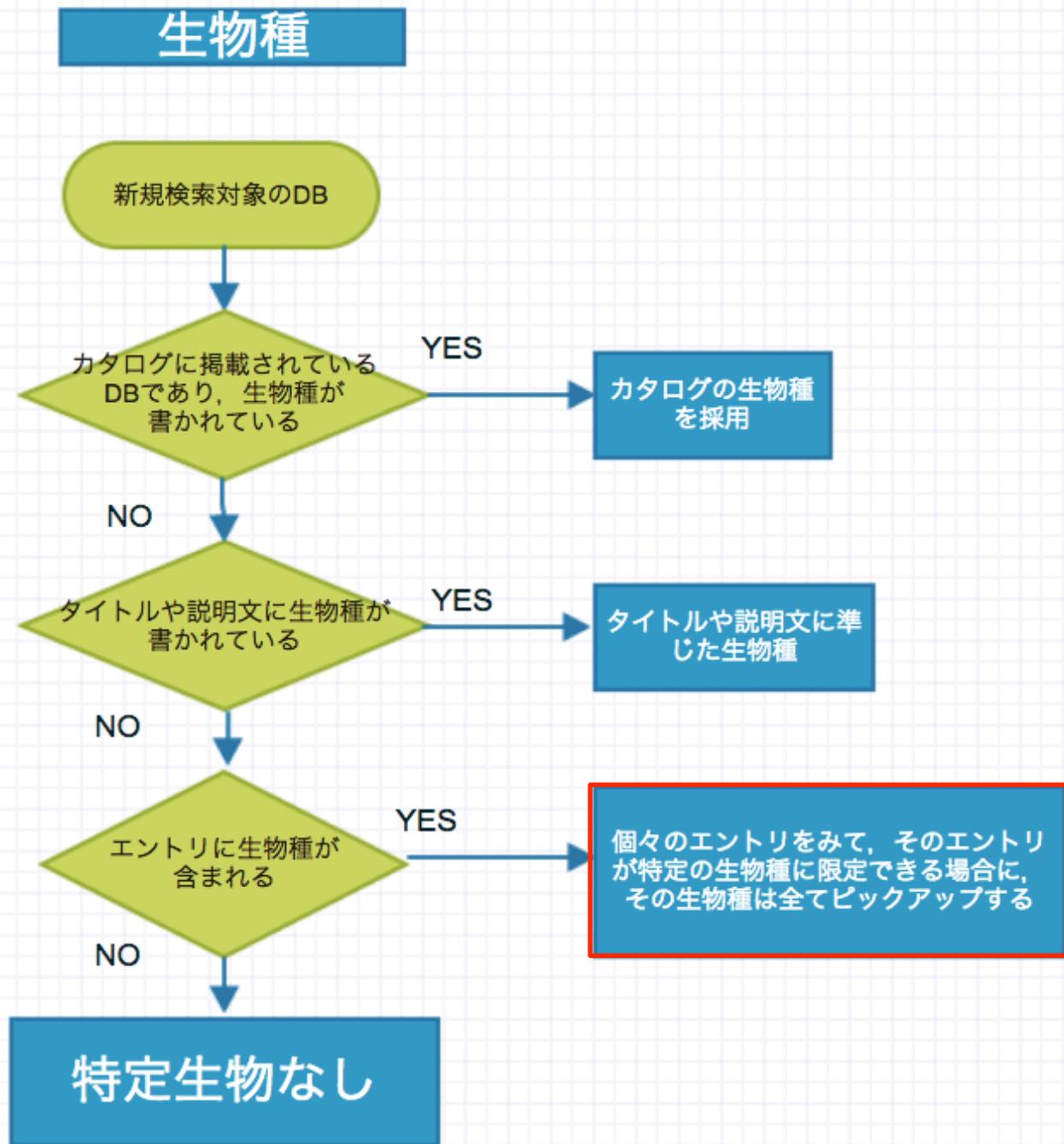
ファセット分類

データベースの種類



ファセット分類

- 全て
- 代表的
- 特定生物なし



ファセット分類

生体の階層

新規検索対象のDB

カタログに掲載されている
DBであり、対応する生体の階層
がカテゴリにある

YES

カタログの分類を
採用

NO

タイトルや説明文に対応する
生体の階層が
書かれている

YES

タイトルや説明文に準じた
生体の階層を適用

NO

エントリーに生体の階層に該当
するデータが含まれる

YES

個々のエントリーをみて、そのエントリー
が特定の生体の階層に限定できる場合に、
その階層は全てピックアップする

NO

その他

- 全て
- 代表的
- その他

Sagaceに追加するDBについて

	National BioResource Project Human embryonic stem cells (ヒトES細胞プロジェクト情報公開ページ)
創薬・疾患に直接関わるDBである。	0
文書系のDBである。	1
独自のデータを公開している。	0
研究者向けであり、一般向けではない。	1
メタデータの取り組みに協力している。	0
計	2
重み	B

DB分類:その他のデータベース
生物種:ヒト
生体の階層:細胞・組織

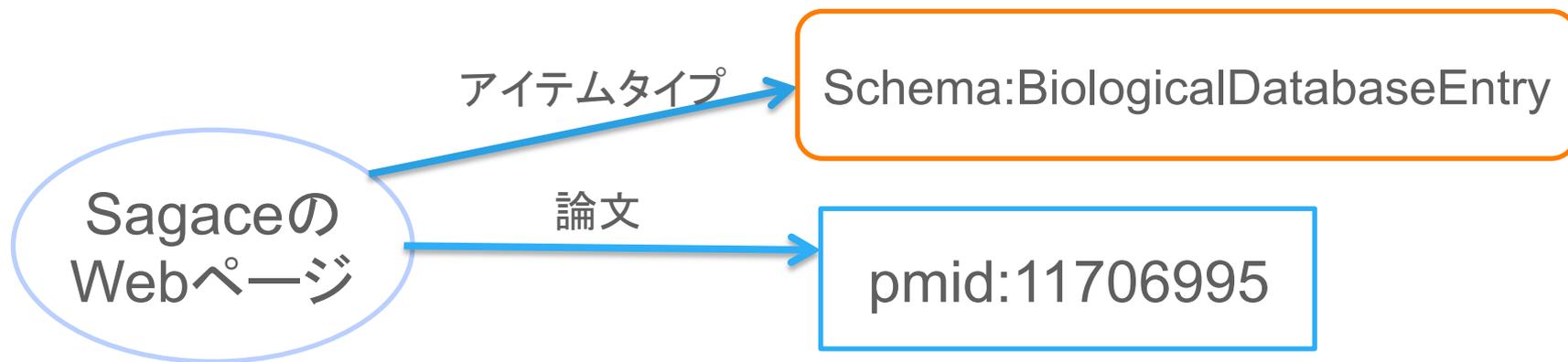
Sagaceに独自に追加するDB について

- <http://sagace.nibio.go.jp/press/members/sagace-candidate-db/>

HTMLとLinked Data

- ・ Html上でタグにプロパティを挿入して繋がる（グラフ）データを作る仕組み。

```
#microdata
<div itemscope="" itemType="http://schema.org/BiologicalDatabaseEntry">
  <meta itemprop='reference' content='pmid:23110816' />
  <a href="http://www.ncbi.nlm.nih.gov/pubmed/23110816">23110816</a>
</div>
```



RDFa Lite

- ・ Microdataと互換性がある。

```
#RDFa Lite
```

```
<div vocab="http://schema.org/" typeof="BiologicalDatabaseEntry">  
  <meta property='reference' content='pmid:11706995' />  
  <a href="http://www.ncbi.nlm.nih.gov/pubmed/11706995">11706995</a>  
</div>
```

```
#microdata
```

```
<div itemscope="" itemtype="http://schema.org/BiologicalDatabaseEntry">  
  <meta itemprop='reference' content='pmid:11706995' />  
  <a href="http://www.ncbi.nlm.nih.gov/pubmed/11706995">11706995</a>  
</div>
```

RDFa Liteのメリット

- ・ 外部の語彙を導入しやすい。

```
<div prefix="fben: http://rdf.freebase.com/ns/">  
  This article's texts is suitable for a wide audience, with a Fog index of  
  <span property="fben:gunning_fog_index">10.2</span>  
</div>
```

- ・ Microdataはschema.orgで定義されている語彙のみ
- ・ RDFa Liteは，RDFで汎用的に使われている語彙も使える。

RDFa Liteのメリット

- Resource属性

```
<p property="author" resource="#mizuki.morita" typeof="Person">  
  <span property="name">Mizuki Morita</span>, <span property="jobTitle">Partner</span>,  
  <span property="worksFor">Aist</span>. Jul 2013.  
</p>
```

- データそれぞれの関係性を記述しやすくなる（参照：デモ）

進捗報告—RDF化

- ・ 希少疾病用医薬品のデータのRDF化
 - ほぼ完了
 - 日本語版の作成
 - IDの付与についてダイナコムと相談

今後の予定

- ・ 8月 & 9月
 - 文献データを検索結果に反映
 - Apache Solrの検討と実装
- ・ 10月
 - トーゴの日
- ・ 11月
 - 実験動物バンクのRDF化
 - 外部データベースとの連携
- ・ 12月
 - 分子生物学会

次回

- ・ 9月20日 (金) or 9月27日 (金) ?