PubAnnotationとPubDictionariesで行う文献アノテーション

金進東　　王 悦
ライフサイエンス統合データベースセンター (DBCLS)

PubDictionaries

- 辞書リポジトリ
  - 誰でも辞書をuploadし、即時文献アノテーションができる。
  - アノテーション速度：20 PubMed abstracts / 秒

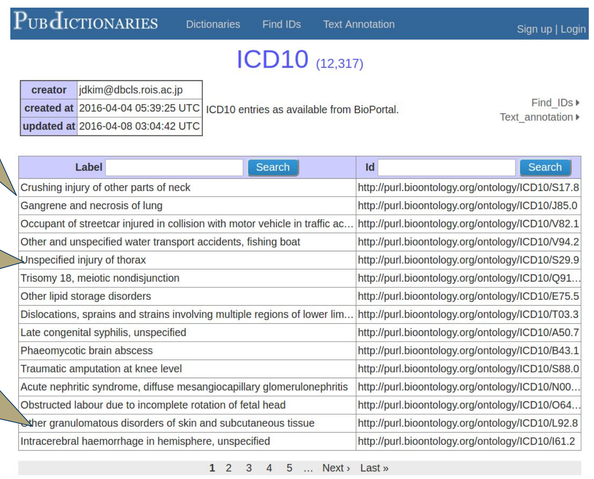
PubAnnotation

- 文献アノテーションリポジトリ
  - 誰でも文献アノテーションを作るのができる。
  - 既に登録されているアノテーションと合わせて使うのができる。

Dictionary = a collection of pairs, (Label, ID)

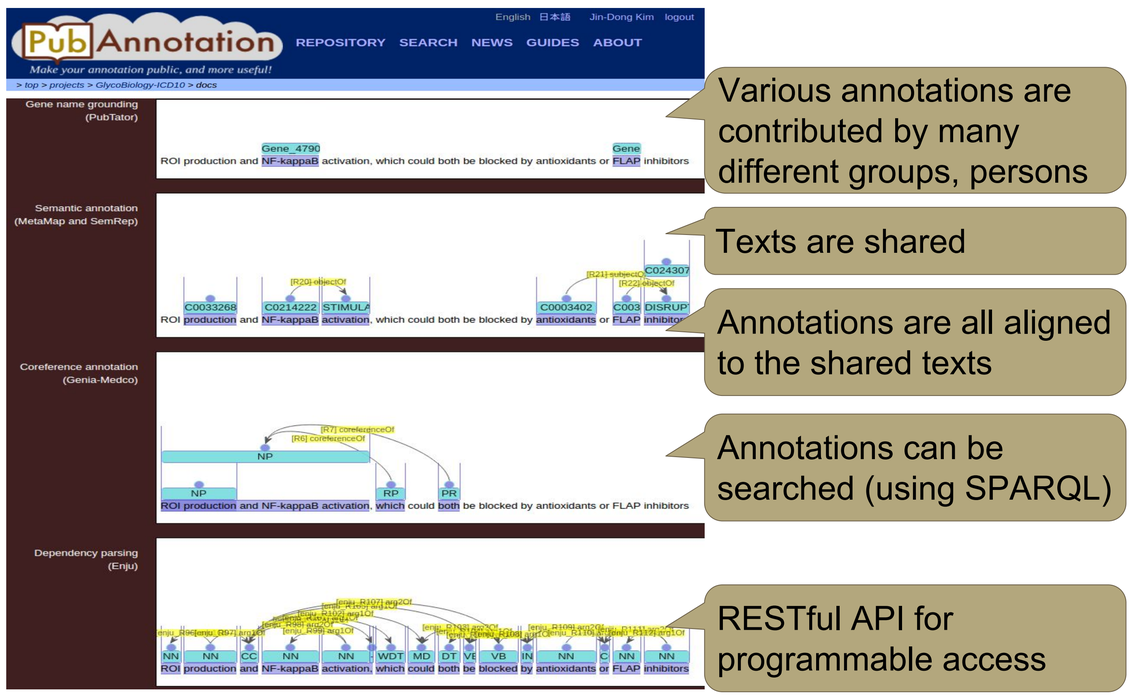
Anyone can upload a Dic. in TSV

It immediately becomes ready for action



Text

Annotation



Various annotations are contributed by many different groups, persons

Texts are shared

Annotations are all aligned to the shared texts

Annotations can be searched (using SPARQL)

RESTful API for programmable access

例: Preeclampsia

Preeclampsia / Colorado大学

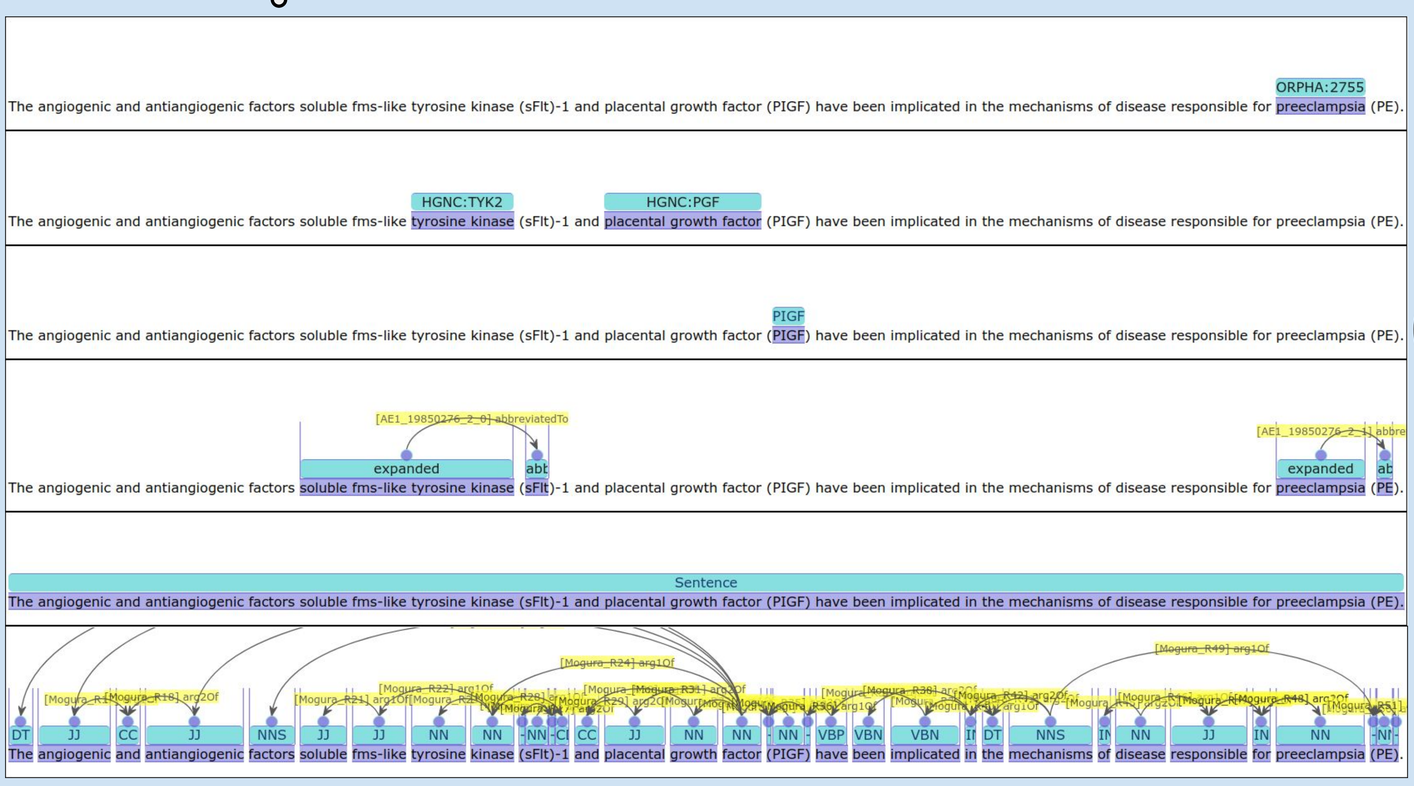
Genes / Colorado大学

GlycoGenes / 創価大学

Allie (Abbreviations) / DBCLS

Sentences / DBCLS

Syntactic Parse / NII, DBCLS



違うグループが違う目的で作ったアノテーションだが、合わせて使える。

検索 (SPARQL)

```
PREFIX pubann:<http://pubannotation.org/ontology/>
PREFIX tao:<http://pubannotation.org/ontology/tao.owl#>
PREFIX prj:<http://pubannotation.org/projects/>
select distinct ?s1 ?s2 where {
    graph prj:preelampsia-gene {
        ?o1 tao:denoted_by ?s1 .
    }

    graph prj:collection-preeclampsia {
        ?o2 tao:denoted_by ?s2 .
    }

    ?s1 tao:belongs_to ?a1 .
    ?s2 tao:belongs_to ?a1 .
}
```

同じ文書で共起するpreeclampsiaと遺伝子のペア

```
PREFIX pubann:<http://pubannotation.org/ontology/>
PREFIX tao:<http://pubannotation.org/ontology/tao.owl#>
PREFIX prj:<http://pubannotation.org/projects/>
select distinct ?s1 ?s2 where {
    graph prj:preelampsia-gene {
        ?o1 tao:denoted_by ?s1 .
    }

    graph prj:collection-preeclampsia {
        ?o2 tao:denoted_by ?s2 .
    }

    ?o3 tao:denoted_by ?s3 .
    ?o3 a pubann:Sentence .

    ?s1 tao:contained_in ?s3 .
    ?s2 tao:contained_in ?s3 .
}
```

同じ文章で共起するpreeclampsiaと遺伝子のペア

```
PREFIX tao:<http://pubannotation.org/ontology/tao.owl#>
PREFIX prj:<http://pubannotation.org/projects/>
PREFIX enju:<http://kmcs.nii.ac.jp/enju/>
select ?s1 ?s2 where {
    graph prj:preelampsia-gene {
        ?o1 tao:denoted_by ?s1 .
    }

    graph prj:collection-preeclampsia {
        ?o2 tao:denoted_by ?s2 .
    }

    ?w1 tao:denoted_by ?s1 .
    ?w2 tao:denoted_by ?s2 .
    ?w1 enju:argOf+ ?w2 .
}
```

構文解析の依存関係で繋がっているpreeclampsiaと遺伝子のペア

CSV 辞書だけ用意すれば　PubAnnotation 100万件のPubMedアブストラクトを　PubDictionaries 一晩でアノテーションできる。

Licensed under a Creative Commons表示4.0国際 (c)2017 金進東（ライフサイエンス統合データベースセンター）