数据库原理项目报告

龙伟淼　程子健　魏嘉璐

2023 年 1 月 19 日

摘要

本次我们小组主要利用 BioGRID（https://downloads.thebiogrid.org/File/BioGRID/Other-Datasets/InteractionToPTMMigration-3.4.125.xlsx）的数据，其中包括了'Interaction ID', 'BioGRID Interactor ID', 'Official Symbol', 'Organism' 是我们需要的列。由于考虑到我们自己的关系表需要满足第三范式（3NF），我们设计了两个查询表，分别：

- Organism{Protein_ID, Protein_Name, Organism_Name}
- Interaction{Protein_ID_A, Protein_ID_B, Interaction_ID}

我们的网站的功能有：

1. 提供基于蛋白质名称的查询，输出蛋白相互作用表
2. 查询某个物种，输出该物种中的所有蛋白-蛋白相互作用

我们小组分工为：

- 魏嘉璐：前端与前后端交互
- 程子健：数据库建立与前后端交互
- 龙伟淼：报告撰写以及整合

1　介绍

The protein–protein interaction data can be represented as a network whose nodes are proteins, and they are connected by edges if the corresponding proteins interact. Previous studies have shown that these networks are highly heterogeneous, containing both a large number of proteins with few interaction partners, but also many highly connected 'hub' proteins. In our project, we manage to extract information out of the protein-protein interaction data and provide users with queries. [1]

2　查询过程

2.1　查询蛋白质

当用户查询蛋白质的名称 p 时，我们的 MySQL 表达式为：

```
01 |  select Interaction_interaction.Interaction_ID, Interaction_organism.
       Organism_Name, Interaction_organism.Protein_Name
02 |  from Interaction_interaction join Interaction_organism
03 |  where Interaction_interaction.Protein_ID_B = Interaction_organism.Protein_ID AND
        Protein_ID_A in (select Protein_ID from Interaction_organism where
       Protein_Name = p)
```

2.2 查询物种

当用户查询物种的名称 o 时，我们的 MySQL 表达式为：

```
select A.Interaction_ID, B.Protein_Name, C.Protein_Name
from Interaction_interaction as A, Interaction_organism as B,
    Interaction_organism as C
where A.Protein_ID_A = B.Protein_ID AND A.Protein_ID_B = C.Protein_ID AND B.
    Organism_Name = o
```

我们小组的编程文件上传在 github 上。(https://github.com/Brycealong/dbsproject)

3 证明第三范式

A relation schema R is in third normal form with respect to a set F of functional dependencies if, for all functional dependencies in F^+ of the form $\alpha \to \beta$, where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:

- $\alpha \to \beta$ is a trivial functional dependency.
- α is a superkey for R.
- Each attribute A in $\beta - \alpha$ is contained in a candidate key for R.

Now, let us again consider the schema for the *Organism* relation, which has the following functional dependencies:

$$Protein\_ID \to Protein\_Name$$
$$Protein\_ID \to Organism\_Name$$

where $Protein\_ID$ is a superkey for *Organism*. The 3NF holds.

Now, let us again consider the schema for the *Interaction* relation, which has the following functional dependencies:

$$Interaction\_ID \to Protein\_ID\_A, Protein\_ID\_B$$

where $Interaction\_ID$ is a superkey for *Interaction*. The 3NF holds.

4 网页展示

通过 html、css 和 js 等工具，我们创建了网页如下：

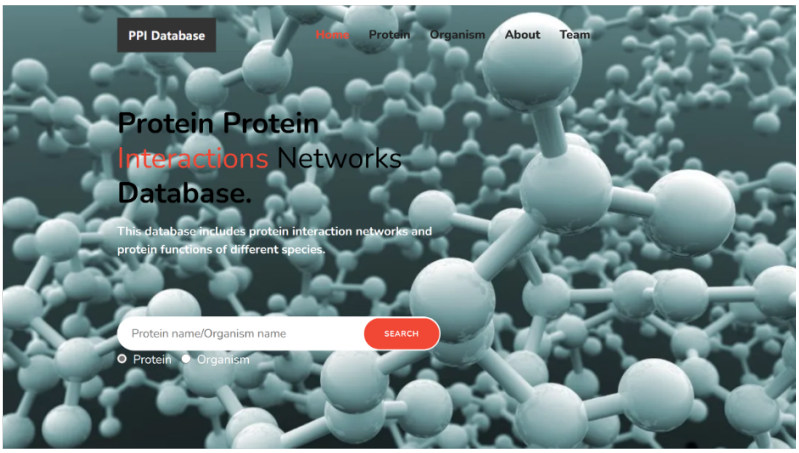

图 1: 网页

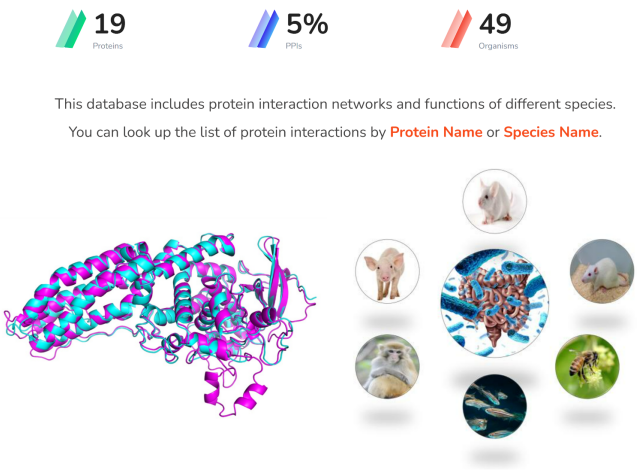
Our Database includes
19
Proteins
5%
PPIs
49
Organisms
This database includes protein interaction networks and functions of different species.
You can look up the list of protein interactions by Protein Name or Species Name.

图 2: 网页 2

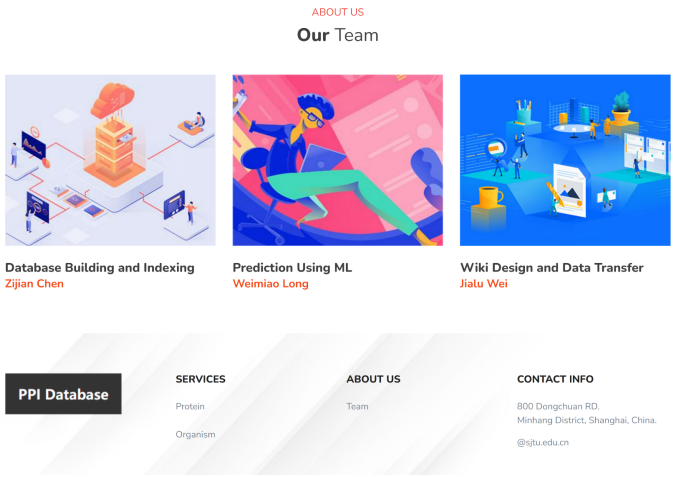
ABOUT US
Our Team
Database Building and Indexing
Zijian Chen
Prediction Using ML
Weimiao Long
Wiki Design and Data Transfer
Jialu Wei
PPI Database
SERVICES
Protein
Organism
ABOUT US
Team
CONTACT INFO
800 Dongchuan RD.
Minhang District, Shanghai, China.
@sjtu.edu.cn

图 3: 网页 3

4.1 查询蛋白质

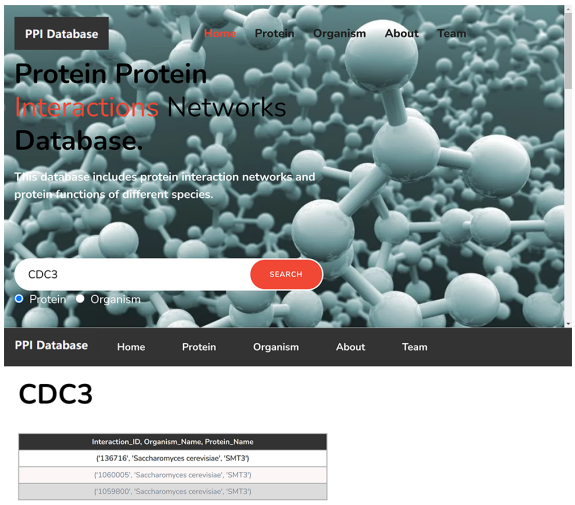

图 4: 查询蛋白质及结果

4.2 查询物种

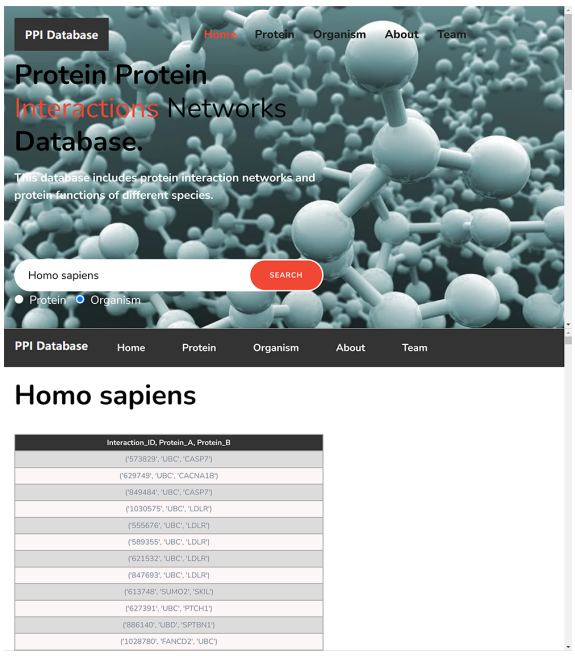

图 5: 查询物种及结果

5 鸣谢

感谢本学期老师吴茂英的辛勤付出，为我们项目提供了无限的灵感。

参考文献

[1] I. Albert and R. Albert, “Conserved network motifs allow protein–protein interaction prediction,” *Bioinformatics*, vol. 20, pp. 3346–3352, 07 2004.