

HIAE Teste-Bioinfo - Instruções [Gianluca Major]

gianlucamajor@gmail.com

Banco de dados

postgresql-9.4

É necessário que exista uma base de dados chamada **hiae-bioinfo**, usuario: **gianluca**, senha: **hiae**

Script para criar tabelas: hiae-python\db\create_tables_on_database.sql

Python

Dependências:

python 2.7

psycopg2

flask

flask_restful

Crawler e API

diretório: /hiae-python

Crawler

executar: `python crawler-get-data-hp.py`

OBJETIVO DA CRAWLER: buscar os dados de fenótipos do link fornecido no teste, tratar e armazenar os dados utilizando Postgre.

- I. Caso seja necessário alterar as informações de acesso ao banco, tais como nome do banco, usuário e senha, é nesse arquivo que essas informações devem ser alteradas para o crawler.
- II. O carregamento do arquivo para o banco pode demorar um pouco. No final do arquivo `crawler-get-data-hp.py`, existe uma linha que pode ser descomentada para exibir o andamento do loading no stdout.
- III. O crawler informa o final do carregamento na base de dados.

API

executar: `python hiae-bioinfo-api.py`

A execução desse comando deve gerar uma saída próxima ao quadro seguinte:

```
* Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)
* Restarting with stat
* Debugger is active!
* Debugger PIN: 527-883-546
```

OBJETIVO DA API: Retornar o(s) fenótipo(s) de um determinado gene ou lista de genes

Exemplos de requisição

Para determinado gene:

`http://localhost:5000/phenotypes/genes/['GLI3']`

Para lista de genes:

`http://localhost:5000/phenotypes/genes/['GLI3','ATR','SNRPB']`

Exemplo de respostas

De sucesso: status code 200:

```
[
  {
    "gene": "ATR",
    "gene_id": 545,
    "phenotype": "Vascular skin abnormality",
    "pho": "HP:0011276"
  },
  {
    "gene": "ATR",
    "gene_id": 545,
    "phenotype": "Weight loss",
    "pho": "HP:0001824"
  }
]
```

De erro: status code 400

```
{
  "Something is wrong in request": "Verify the url, parameters and format."
}
```

De erro: status code 500 (falha de conexão com o banco)

```
{  
  "Ops!": "We are having a problem, please come back later."  
}
```

Disponibilidade Online

Disponibilizei a API em python em uma instância da amazon. Não tenho um domínio próprio, desta forma as urls estão utilizando ip real. Por consequência de uma possível *reboot* da instância, esse ip pode mudar, por isso, se tiver problema para acessar, não hesite em me contactar.

Exemplos de requisição

Para determinado gene:

[http://54.233.236.135:5000/phenotypes/genes/\['GLI3'\]](http://54.233.236.135:5000/phenotypes/genes/['GLI3'])

Para lista de genes:

[http://54.233.236.135:5000/phenotypes/genes/\['GLI3','ATR','SNRPB'\]](http://54.233.236.135:5000/phenotypes/genes/['GLI3','ATR','SNRPB'])

Java

Dependências:

Java jre 1.8 ou mais nova
apache-tomcat-8.5

API

diretório: /hiae-java/bioinfo/

A forma mais simples de executar o projeto java localmente é baixando o tomcat já com o war do teste-bioinfo. Depois de baixar [deste link](https://drive.google.com/open?id=0B1LDtC9CW06gYi1ZNTZua3pDOW8) (https://drive.google.com/open?id=0B1LDtC9CW06gYi1ZNTZua3pDOW8), basta executar `apache-tomcat-8.5.16/bin/catalina.sh run`; uma vez que o tomcat estiver rodando, deve ser possível executar as seguintes requisições:

Exemplos de requisição

OBJETIVO DA API: Retornar o(s) fenótipo(s) de um determinado gene ou lista de genes

Para determinado gene:

`http://localhost:8080/bioinfo/phenotypes/genes/['GLI3']`

Para lista de genes:

`http://localhost:8080/bioinfo/phenotypes/genes/['GLI3','ATR','SNRPB']`

Disponibilidade Online

Também disponibilizei a API em Java na amazon. Assim como no caso da api de python, não tenho um domínio próprio, desta forma as urls estão utilizando ip real. Por consequência de uma possível *reboot* da instância, esse ip pode mudar, por isso, se tiver problema para acessar, não hesite em me contactar.

Para determinado gene:

`http://54.233.236.135:8080/bioinfo/phenotypes/genes/['GLI3']`

Para lista de genes:

`http://54.233.236.135:8080/bioinfo/phenotypes/genes/['GLI3','ATR','SNRPB']`

Apenas se for compilar local:

Dependências:

Java jdk version 1.8 ou mais nova
maven

API

diretório: /hiae-java/bioinfo/

executar o comando: mvn package

O resultado deve ser algo parecido com o apresentado abaixo:

```
[INFO] Building war: /home/gianluca/hiae-java/bioinfo/target/bioinfo.war
[INFO] WEB-INF/web.xml already added, skipping
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 5.870 s
[INFO] Finished at: 2017-07-14T18:19:42+00:00
[INFO] Final Memory: 18M/45M
[INFO] -----
```

A primeira vez que o package for executado ele baixará todas as dependências do projeto.

Depois package, deve existir um diretório /target e dentro dele um arquivo chamado bioinfo.war. /target/bioinfo.war

O arquivo /bioinfo.war deve ser copiado/movido para o diretório de webapps do tomcat.

Deve ficar algo assim: /apache-tomcat-8.5.16/webapps/bioinfo.war

Agora basta executar o tomcat:

apache-tomcat-8.5.16/bin/catalina.sh run; uma vez que o tomcat estiver rodando, deve ser possível executar as seguintes requisições:

Exemplos de requisição

OBJETIVO DA API: Retornar o(s) fenótipo(s) de um determinado gene ou lista de genes

Para determinado gene:

[http://localhost:8080/bioinfo/phenotypes/genes/\['GLI3'\]](http://localhost:8080/bioinfo/phenotypes/genes/['GLI3'])

Para lista de genes:

[http://localhost:8080/bioinfo/phenotypes/genes/\['GLI3','ATR','SNRPB'\]](http://localhost:8080/bioinfo/phenotypes/genes/['GLI3','ATR','SNRPB'])

