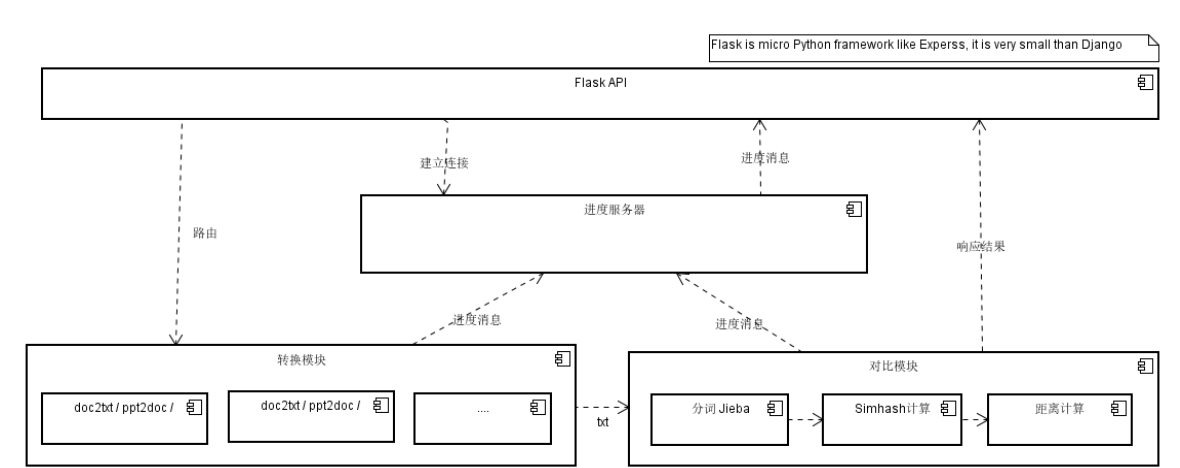# 班级内文章对比查重系统

使用 simihash 算法, 对班级内的文章进行查重对比。

系统架构

核心算法

simhash

How to test

```
python test.py
```

Docker

Repo

registry.cn-chengdu.aliyuncs.com/sunwu/arktao.textcompare:v0.1.11

run

```
docker run  -p 8217:8000  -itd --restart always  --privileged  --name
arktao.textcomparev0.1.11 registry.cn-
chengdu.aliyuncs.com/sunwu/arktao.textcompare:v0.1.11
```

RestFul API

host

```
122.114.95.64:8217
```

文章对比

POST /v1/html

req

```json
{
  "type": "html",
  "distance":20,
  "uid":"sunwutest",   //用于获取识别进度
  "papers": [
    {
      "name": "zhangsan",
      "sno": "2019201964",
      "papername": "if i was a boy",
      "content":"<h1>if i was a boy, i will go to fly like a bird</h1>"
    },
      {
      "name": "zhangsan",
      "sno": "2019201963",
      "papername": "if i was a boy",
      "content":"<h1>if i was a boy, i will go to fly like a bird</h1>"
    },
      {
      "name": "zhangsan",
      "sno": "2019201962",
      "papername": "if i was a boy",
      "content":"<h1>if i was a boy, i will go to fly like a bird</h1>"
    }
  ]
}
```

res

```json
{
  "code": "成功",
  "data": {
    "2019201961": [
      {
        "distance": -1,
        "name": "zhangsan",
        "papername": "if i was a boysss",
        "sno": "2019201961"
      },
      {
        "distance": 8,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "percent": 77.77777777777779,
        "sno": "2019201963"
```

```json
      },
      {
        "distance": 8,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "percent": 77.77777777779,
        "sno": "2019201964"
      },
      {
        "distance": 8,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "percent": 77.77777777779,
        "sno": "2019201962"
      }
    ],
    "2019201962": [
      {
        "distance": -1,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "sno": "2019201962"
      },
      {
        "distance": 0,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "percent": 100.0,
        "sno": "2019201963"
      },
      {
        "distance": 0,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "percent": 100.0,
        "sno": "2019201964"
      },
      {
        "distance": 8,
        "name": "zhangsan",
        "papername": "if i was a boysss",
        "percent": 77.77777777779,
        "sno": "2019201961"
      }
    ],
    "2019201963": [
      {
        "distance": -1,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "sno": "2019201963"
      },
      {
        "distance": 0,
        "name": "zhangsan",
        "papername": "if i was a boy",
        "percent": 100.0,
        "sno": "2019201964"
```

```
        },
        {
          "distance": 0,
          "name": "zhangsan",
          "papername": "if i was a boy",
          "percent": 100.0,
          "sno": "2019201962"
        },
        {
          "distance": 8,
          "name": "zhangsan",
          "papername": "if i was a boysss",
          "percent": 77.77777777777779,
          "sno": "2019201961"
        }
      ],
      "2019201964": [
        {
          "distance": -1,
          "name": "zhangsan",
          "papername": "if i was a boy",
          "sno": "2019201964"
        },
        {
          "distance": 0,
          "name": "zhangsan",
          "papername": "if i was a boy",
          "percent": 100.0,
          "sno": "2019201963"
        },
        {
          "distance": 0,
          "name": "zhangsan",
          "papername": "if i was a boy",
          "percent": 100.0,
          "sno": "2019201962"
        },
        {
          "distance": 8,
          "name": "zhangsan",
          "papername": "if i was a boysss",
          "percent": 77.77777777777779,
          "sno": "2019201961"
        }
      ]
    },
    "message": "成功",
    "status": 200,
    "type": "sshd"
  }
```

note

```
  "type": "html", // 文章类型
  "distance":20,  //判定距离
  "papers": [      //文章
```

获取进度

连接进度服务器

进度服务器用户，先使用指定用户名连接进度服务器，用于接受进度。如下 uid 表示用户的的名称，接受进度的时候要用到。

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <title>Title</title>
</head>
<body>
<script src='http://cdn.bootcss.com/socket.io/1.3.7/socket.io.js'></script>
<script>

    // 初始化io对象
    var socket = io('wss://ppt2img_socket.lessonplan.cn:8202');

    // uid 用户的uid
    var uid = "sunwutest";

    //发送连接进度服务器请求
    socket.on('connect', function () {
            socket.emit('connect_progress', uid);
        }
    );

    //服务器返回进度信息,这里不返回图片信息
    socket.on('progress_msg', function (msg) {
        console.log(JSON.parse(msg));
    });
</script>
</body>
</html>
```

进度内容

```
{
    "msg":"进度消息",
    "code": 0, // 状态码, -1：错误，0：正确
    "progress": 100.0 //进度消息，有效范围[0,100]
 }
```

已配置

服务自动恢复

宕机时间最多一分钟,服务挂了之后会自启