**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN - ĐHQG TPHCM**

**KHOA CÔNG NGHỆ PHẦN MỀM**

**ĐỒ ÁN MÔN HỌC CHUYÊN NGÀNH**

ĐỀ TÀI: HỆ THỐNG KHUYẾN NGHỊ VIỆC LÀM

*GVHD:*

*TS HUỲNH NGỌC TÍN*

*SINH VIÊN THỰC HIỆN:*

*12520248 - Trần Minh Luận*

*12520492 – Nguyễn Thanh Anh Tuyên*

MỤC LỤC

[**GIỚI THIỆU BÀI TOÁN** 1](#_Toc453677114)

[**1.** **XÁC ĐỊNH VÀ ĐẶC TẢ YÊU CẦU** 2](#_Toc453677115)

[**1.1.** **DANH SÁCH CÁC YÊU CẦU** 2](#_Toc453677116)

[1.1.1. Phân tích, thống kê dữ liệu đầu vào. 2](#_Toc453677117)

[1.1.2. Cho phép chạy thực nghiệm các thuật toán khuyến nghị. 2](#_Toc453677118)

[1.1.3. So sánh các thuật toán khuyến nghị 2](#_Toc453677119)

[1.1.4. Cung cấp API khuyến nghị 2](#_Toc453677120)

[**1.2.** **MÔ TẢ CHI TIẾT TỪNG YÊU CẦU** 2](#_Toc453677121)

[1.2.1. Phân tích, thống kê dữ liệu đầu vào. 2](#_Toc453677122)

[1.2.2. Cho phép chạy thực nghiệm các thuật toán khuyến nghị. 2](#_Toc453677123)

[1.2.3. So sánh các thuật toán khuyến nghị. 2](#_Toc453677124)

[1.2.4. Cung cấp API khuyến nghị 2](#_Toc453677125)

[**2.** **THIẾT KẾ** 3](#_Toc453677126)

[**2.1.** **KIẾN TRÚC HỆ THỐNG** 3](#_Toc453677127)

[2.1.1. Kiến trúc 3](#_Toc453677128)

[2.1.2. Mô tả kiến trúc 4](#_Toc453677129)

[**2.2.** **THIẾT KẾ DỮ LIỆU** 4](#_Toc453677130)

[2.2.1. System DB 4](#_Toc453677131)

[2.2.2. User dataset 6](#_Toc453677132)

[**2.3.** **THIẾT KẾ XỬ LÝ** 6](#_Toc453677133)

[2.4. **THIẾT KẾ GIAO DIỆN** 6](#_Toc453677134)

[**3.** **LẬP TRÌNH VÀ KIỂM THỬ** 7](#_Toc453677135)

[**KẾT LUẬN** 8](#_Toc453677136)

# **GIỚI THIỆU BÀI TOÁN**

Hệ thống “Đánh giá các thuật toán khuyến nghị” (Recommender algorithm evaluation) là hệ thống cho phép người dùng đưa vào các tập dữ liệu thực nghiệm để tiến hành các phân tích, thống kê hay các thực nghiệm và đánh giá dựa trên tập dữ liệu đầu vào. Hệ thống này phục vụ cho việc giải quyết bài toán khuyến nghị việc làm bằng cách cung cấp các đánh giá, so sánh các kết quả thực nghiệm. Từ đó giúp cho việc giải quyết bài toán khuyến nghị việc làm dễ dàng hơn.

# **XÁC ĐỊNH VÀ ĐẶC TẢ YÊU CẦU**

## **DANH SÁCH CÁC YÊU CẦU**

### Xem thống kê dữ liệu đầu vào:

* Người dùng sẽ chọn tập dữ liệu cần phân tích để đưa vào module phân tích dữ liệu. Nếu chưa có thì người dùng có thể upload lên hệ thống. Tập dữ liệu thực hiện gồm có 3 file.
  + File Cv.txt: Chứa dữ liệu về người dùng bao gồm các thông tin về cv người dùng.
  + File Job.txt: Chứa dữ liệu về việc làm sẽ được khuyến nghị.
  + Score.txt: Chứa dữ liệu về sở thích của người dùng trong file Cv.txt đối với những việc làm trong file Job.txt
* Sau khi người dùng chọn dữ liệu đầu vào. Người dùng chọn chức năng phân tích để bắt đầu phân tích dữ liệu. Sau khi hệ thống phân tích xong sẽ hiển thị ra màn hình cho người dùng.

### Cho phép chạy thực nghiệm các thuật toán khuyến nghị.

* Người dùng tạo chọn dataset cần thực nghiệm, chọn thuật toán cần chạy. Nếu tập dữ liệu chưa có thì người dùng có thể upload lên hệ thống. Sau đó người dùng bấm chạy. Hệ thống sẽ thực hiện chạy thuật toán khuyến nghị. Sau khi chạy xong. Hệ thống sẽ thông báo cho người dùng kết quả thực nghiệm và lưu trữ lại vào file output.txt.
* Cấu trúc file output gồm 3 cột: UserId, JobId, Score

### So sánh các thuật toán khuyến nghị

* Người dùng sẽ nhập những file output.txt sinh ra khi chạy thực nghiệm vào hệ thống. Hệ thống sẽ thực hiện tính toán và trả về kết quả kết quả so sánh các thuật toán khuyến nghị cho người dùng.

## **PHÂN TÍCH YÊU CẦU**

### Biểu đồ phân rã chức năng hệ thống

Recomender evalution system

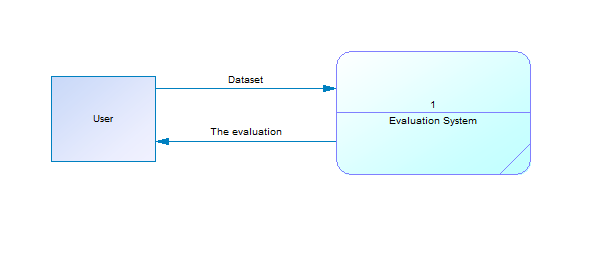
Upload dataset

View dataset Statistic

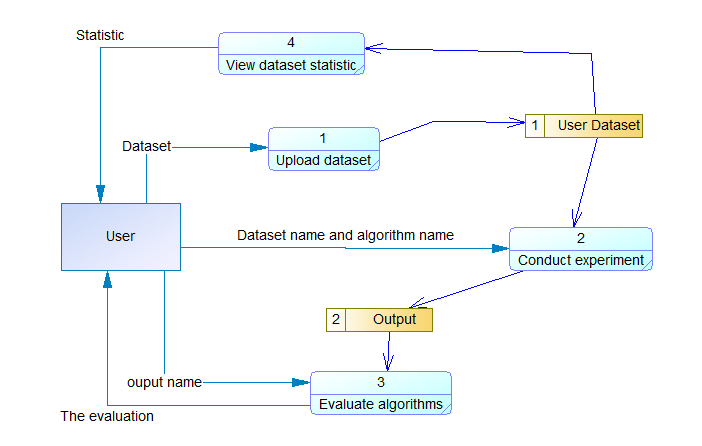
Conduct experiment

Evaluate algorithms

### Sơ đồ DFD mức 0



### Sơ đồ DFD mức 1



# **THIẾT KẾ**

## **KIẾN TRÚC HỆ THỐNG**

### Kiến trúc

**SYSTEM DB**

**User DataSet**

**Spring MVC**

**Apache Mahout**

**Apache Hadoop**

**Recommender web app**

**Recommender webservice**

**Algorithms**

**Evaluation metrics**

**Visualization**

**Experimental conduction**

**Data preparation**

**System Interfaces**

**System core modules**

**Libraries**

### Mô tả kiến trúc

* + System interfaces
* Recommender web app: Cung cấp giao diện để người dùng có thể upload dữ liệu của họ. Sau đó có thể tiến hành thống kê hay chạy các thuật toán khuyến nghị.
* Recommender web service: Cung cấp một số API để người dùng có thể nhúng các kết quả khuyến nghị vào hệ thống riêng của họ.
  + System core modules
* Visualization: Đây là module cho phép xuất các kết quả thống kê, kết quả thực hiện của các thuật toán khuyến nghị,…
* Evaluation metrics: thực hiện các độ đo giữa các thuật toán khuyến nghị
* Experimental conduction: chạy thực nghiệm các thuật toán khuyến nghị.
* Algorithms: Chứa các thuật toán khuyến nghị như CF, CB, Hybrid,…
* Data preparation: Dùng để chuẩn bị dữ liệu đầu vào cho các thuật toán khuyến nghị. Thực hiện thống kê dữ liệu trên các đặc trưng.
  + Database
* System DB
* User dataset
  + Libraries
* Spring MVC
* Mahout
* Hadoop

## **THIẾT KẾ DỮ LIỆU**

## System DB

Database này gồm có 2 bảng sau:

* USER (UserId, UserName, Email, Password)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STT | Tên cột | Mô tả | Kiểu dữ liệu | Miền giá trị | Ghi chú |
| 1 | UserId | Mã người dùng | Integer | 11 | Not null, primary key |
| 2 | UserName | Tên người dùng | Nvarchar | 50 | Not null |
| 3 | Email | Email đăng nhập người dùng | Varchar | 50 | Not null |
| 4 | password | Mật khẩu đăng nhập | Varchar | 50 | Not null |

* Task (TaskId, UserId, TaskName, TimeCreate, Status, Algorithm, InputFolder, OutputFile)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STT | Tên cột | Mô tả | Kiểu dữ liệu | Miền giá trị | Ghi chú |
| 1 | TaskId | Mã của task | Integer | 11 | Not null, primary key |
| 2 | UserId | Mã người dùng | Integer | 11 | Not null, foreign key đến User |
| 3 | TaskName | Tên task | Nvarchar | 100 | Not null |
| 4 | TimeCreate | Thời gian tạo task | Datetime |  | Not null |
| 5 | Status | Trạng thái task | Nvarchar | 10 | Not null |
| 6 | Algorithm | Tên thuật toán | Varchar | 20 | Not null |
| 7 | InputFolder | Đường dẫn đến thư mục chứa file input | Varchar | 100 | Not null |
| 8 | OutputFile | Đường dẫn file output | Varchar | 100 | Not null |

## User dataset

* Dữ liệu đầu vào để chạy các thuật toán gồm 3 file Job.txt, CV.txt và Score.txt. Quy ước trong các file .txt mỗi cột cách nhau bằng một ký tự tab.
* File Job.txt gồm các cột sau: JobId, JobName, Location, Salary, Category, Requirement, Tag, Description.
* File CV.txt gồm các cột sau: UserId, CVId, UserName, CVName, UserAddress, ExpectedSalary, Category, Language, Education, Skill, CareerObjective.
* File Score.txt gồm các cột sau: UserId, JobId, Score.
* Dữ liệu đầu ra

File output.txt có cấu trúc gồm 3 cột: UserId, JobId, Score.

## **THIẾT KẾ GIAO DIỆN**

# **LẬP TRÌNH VÀ KIỂM THỬ**

# **KẾT LUẬN**