**JOBRECSYS ARCHITECTURE**

MỤC LỤC

[**Phần 1.** **KIẾN TRÚC TỔNG QUAN** 1](#_Toc456071858)

[1.1. Mô hình kiến trúc 1](#_Toc456071859)

[1.2. Mô tả kiến trúc 2](#_Toc456071860)

[**Phần 2.** **KIẾN TRÚC CHI TIẾT CÁC SUB-PROJECT** 3](#_Toc456071861)

[2.1. Project JobRecAlgComparer 3](#_Toc456071862)

[2.2. Project JobRecTaskManagement 5](#_Toc456071863)

[2.3. Project CreateDataSet 8](#_Toc456071864)

# **KIẾN TRÚC TỔNG QUAN**

## Mô hình kiến trúc

**SYSTEM DB**

**User DataSet**

**Spring MVC**

**Apache Mahout**

**Apache Hadoop**

**JobRec Task Management**

**Dataset Analyzation**

**Experimental conduction & comparision**

**Web module**

**Core**

**System DB**

**Libraries**

## Mô tả kiến trúc

* + Web module
* JobRec Task Management: Mô-đun này cung cấp cho người dùng chức năng quản lý dataset, xem thống kê dataset, quản lý các tiến trình thực thi các thuật toán khuyến nghị, xem kết quả và các đánh giá kết quả khuyến nghị.
  + Core
* Dataset analyzation: Đây là mô-đun phân tích dataset của người dùng.
* Experimental conduction & comparision: chạy thực nghiệm các thuật toán khuyến nghị và so sánh kết quả thực hiện.
  + Database
* System DB: database dùng để tổ chức lưu trữ thông tin người dùng, các tác vụ thực hiện của người dùng và các thông tin thiết lập chạy thuật toán.
* User dataset: hệ thống lưu trữ files dataset do người dùng upload cùng với đó là các file kết quả khuyến nghị, kết quả so sánh và kết quả thống kê.
  + Libraries
* Spring MVC: dùng hỗ trợ xây dựng website tương tác người dùng.
* Mahout: sử dụng các thư viện recommend có sẵn trong Mahout để thực hiện khuyến nghị.
* Hadoop: sử dụng để lưu trữ dữ liệu lớn, hỗ trợ cho việc phát triển trong tương lai.

# **KIẾN TRÚC CHI TIẾT CÁC SUB-PROJECT**

Hệ thống được thiết kế dựa vào mô hình Client-Server. Bởi vì tiến trình thực hiện các thuật toán khuyến nghị, tiến trình phân tích dữ liệu là các tiến trình tốn nhiều thời gian nên không phù hợp cho các ứng dụng web. Chính vì lí do này hệ thống được chia thành nhiều project con. Và các project con này sẽ gọi lẫn nhau và trao đổi thông tin cho nhau. Gồm các sub-project như sau.

Client computers

JobRec Task Management

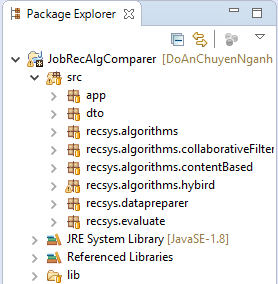
Analyzation

JobRecAlg Comparer

Database

## Project JobRecAlgComparer

Cấu trúc thư mục và package trong project

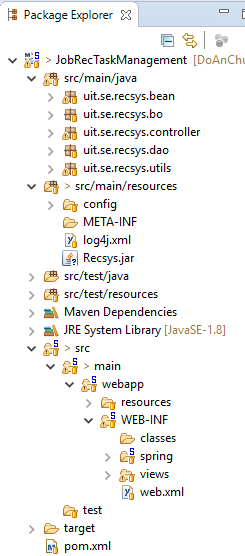


Mô tả các package và folder:

* + src.app: chứa main class thực thi chương trình
  + src.dto: chứa các class dto
  + src.algorithms: chứa 3 package con và 1 abstract class về thuật toán gồm:
    - abstract class RecommendationAlgorithm: class cha chứa các phương thức thiết lập đầu vào và ra của dataset, class này dùng cho các class cài đặt thuật toán kế thừa.
    - src.algorithms.collaborativeFiltering: chứa class cài đặt thuật toán collaborative filtering
    - src.algorithms.contentBased: chứa class cài đặt thuật toán content-based
    - src.algorithms.hybrid: chứa class cài đặt thuật toán hybrid
  + src.datapreparer: chứa các class tiền xử lý dữ liệu trước khi đưa vào cho các thuật toán chạy
  + src.evaluate: chứa các class cài đặt việc so sánh thuật toán.
* Thư mục lib: chứa các thư viện của project

## Project JobRecTaskManagement

Cấu trúc thư mục và package trong project

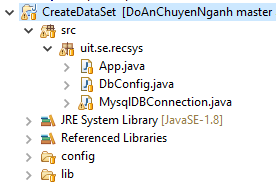


Mô tả các package và folder:

* **Package src/main/java**: là package chính chứa toàn bộ các package trong project gồm:
  + uit.se.recsys.controller: chứa các file controller
  + uit.se.recsys.bean: chứa các file java bean
  + uit.se.recsys.dao: chứa các class java tương tác với database
  + uit.se.recsys.bo: chứa các class xử lý nghiệp vụ
  + uit.se.recsys.utils: chứa các class tiện ích
* **Package src/main/resources**: là package chứa các file tài nguyên của chương trình. Package này chứa folder quan trọng sau:
  + Config: chứa các file config cho hệ thống gồm:
    - File message.properties: là file chứa các thông báo lỗi dùng xuất ra màn hình cho người dùng
    - File webDBConfig.properties: là file chứa các thông tin cấu hình kết nối đến mysql server.
* **Package src/main/webapp**: gồm các file và folder liên quan đến phần cấu hình project và các views trong project:
  + Folder resources: chứa tài nguyên của project sử dụng để hổ trợ lập trình giao diện website:
    - Thư mục css: chứa các file .css
    - Thư mục js: chứa các file .js
    - Thư mục images: chứa các hình ảnh
    - Thư mục libs: chứa các thư viện của các bên thứ 3 hỗ trợ làm giao diện web.
  + Folder WEB-INF: chứa các file cấu hình:
    - Spring/appServlet/servlet-context.xml: cấu hình các bean cho ứng dụng web (web bean)
    - Spring/root-context.xml: cấu hình các bean cho toàn ứng dụng (non-web bean)
    - Views: chứa các file views của web (file .jsp)
    - Web.xml: cấu hình web, là file đầu tiên được nạp vào server khi ứng dụng chạy để cấu hình và nạp các tập tin cấu hình khác.
* **File pom.xml**: là file cấu hình các dependences (thư viện: spring, JDPC, …) của ứng dụng web, các dependences này sẽ được maven tự động tải về thêm vào project khi build project lần đầu.

## Project CreateDataSet

Cấu trúc thư mục và package trong project



Mô tả các package và folder:

* Package uit.se.recsys: chứa các class sau:
  + App.java: class main thực thi chương trình
  + DbConfig.java: class cấu hình kết nối đến database
  + MysqlDBConnection.java: class tạo connection và thực thi các câu lệnh truy vấn sql.
* Thư mục config: chứa file config.txt chứa các thông số cấu hình database cần kết nối
* Thư mục lib: chứa thư viện của project