Mô tả kiến trúc hệ thống:

* Người dùng tương tác với module thu thập tài liệu của hệ thống bằng cách nhập từ khóa vào sau đó yêu cầu hệ thống tìm kiếm. Hệ thống sẽ tìm các bài báo khoa học trên các thư viện số, các bài báo này phải chứa thông tin chỉ mục liên quan đến từ khóa do người dùng đưa ra.
* Sau khi thu thập xong người dùng có thể lưu lại kết quả hoặc chọn chức năng phân lớp theo chủ đề.
* Module phân lớp theo chủ đề dựa vào tiêu đề và tóm tắt của các bài báo vừa được thu thập về hoặc các bài báo trong cơ sở dữ liệu có sẵn. Tạo ra nhóm các bài báo có chủ đề giống nhau.
* Hệ thống kiểm tra các bài báo này xem có trùng lặp với các dữ liệu đã có trong cơ sở dữ liệu không. Nếu trùng thì loại bỏ tài liệu đó ra.
* Cuối cùng lưu các kết quả tìm kiếm xuống cơ sở dữ liệu.

Mô tả module thu thập:

* Khi người dùng đưa ra yêu cầu tìm kiếm các bài báo khoa học theo tên tác giả, tiêu đề hoặc hội nghị. Hệ thống sẽ tìm kiếm thông tin các bài báo liên quan từ ba thư viện số là ACM, Citeseer, IEEExplore.
* Mỗi bài báo tìm được sẽ được rút các thông tin metadata và hiển thị lên cho người dùng. Thông tin metadata gồm tiêu đề, tác giả, hội nghị, năm công bố, tóm tắt, phần tham khảo.

Mô tả module phân lớp:

* Đầu tiên cho hệ thống học từ tập dữ liệu học gồm có các bài báo khoa học và khung phân loại chủ đề thuộc lĩnh vực khoa học máy tính đã nói ở trên.
* Dựa vào những gì đã được học và thông tin tiêu đề, tóm tắt mỗi bài báo trong danh sách các bài báo cần được phân loại để rút ra các đặc trưng cho mỗi bài báo.
* Từ các đặc trưng này sử dụng thuật toán SVM ([http://en.wikipedia.org/wiki/ Support\_vector\_machine](http://en.wikipedia.org/wiki/%20Support_vector_machine) ) để phân loại các bài báo ra theo từng chủ đề khác nhau.

Module import dữ liệu :

* Từ cơ sở dữ liệu các bài báo khoa học có sẵn hệ thống này sẽ rút ra các thông tin metadata như tiêu đề, tác giả, hội nghị, năm công bố, tham khảo.