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 sau đó kết quả trả về được xử lý trong module lấy thông tin.
* Module lấy thông tin sẽ phân tích kết quả trả về để lấy những thông tin về metadata của bài báo từ đó người dùng có thể lưu lại kết quả hoặc chọn chức năng phân lớp chủ đề của bài báo. Hệ thống cho phép chỉnh sửa,bổ xung thông tin cần thiết về dữ liệu vừa rút ra được
* Module phân lớp theo chủ đề dựa vào tiêu đề hoặc tóm tắt của các bài báo vừa được thu thập được xác định chủ đề của bài báo theo khung phân loại của chương trình.
* Khi lưu dữ liệu hệ thống sẽ kiểm tra các bài báo vừa tìm kiếm được 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 (các thông tin cho phép người dùng tìm kiếm được quy định bởi các thư viện số). Hệ thống sẽ tìm kiếm thông tin các bài báo liên quan từ ba thư viện số là ACM, Citeseer, IEEExplore.
* Kết quả trả về từ hệ thống thư viện số là đường dẫn tới các bài báo mà hệ thống tìm được. Những đường link này là đầu vào cho module lấy thông tin các bài báo khao học.

Mô tả module lấy thông tin:

* Mỗi bài báo tìm được sẽ được rút các thông tin metadata bằng các parser 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 lấy ra các thông tin metadata như tiêu đề, tác giả, hội nghị, năm công bố, tham khảo.
* Các bài báo từ dữ liệu cơ sở có sẵn sẽ được xác định chủ đề sau đó thêm vào cơ sở dữ liệu của chương trình.