|  |
| --- |
| BAN CƠ YẾU CHÍNH PHỦ  **HỌC VIỆN KỸ THUẬT MẬT MÃ**    ĐỒ ÁN TỐT NGHIỆP  **XÂY DỰNG VÀ ĐẢM BẢO AN TOÀN CHO HỆ THỐNG THU THẬP VÀ CHIA SẺ TIN TỨC TRỰC TUYẾN**  Ngành: An toàn thông tin  Mã số: 7.48.02.02  *Sinh viên thực hiện*:  **Phạm Hảo Nam**  Lớp: AT13G  *Người hướng dẫn*:  **TS. Trần Anh Tú**  Khoa An toàn thông tin – Học viện Kỹ thuật mật mã    **Hà Nội, tháng 2 năm 2021** |

# PHẦN 1: KHẢO SÁT VÀ PHÂN TÍCH HỆ THỐNG

## I.PHÂN TÍCH YÊU CẦU

### 1.1.1 Chức năng

Đây là hệ thống thu thập dữ liệu và phân tích dữ liệu nhằm mục đích chia sẻ dữ liệu tin cậy và kết hợp với các hệ thống khác tạo thành hệ sinh thái.

* Cho phép user các trang web khác cùng truy thực hiện login và sử dụng tài nguyên chung
* Thu thập và phân tích dữ liệu từ các trang web tin cậy. Từ đó có thể phân tích và chuẩn hóa dữ liệu nhằm phục vụ cho mục đích tìm kiếm nhanh và tiếp cận dễ dàng cho người dung
* Cho phép người dùng bình luận các bài báo hiển thị.
* Hiện thị bài báo dựa trên hành vi người dùng.
* Quản trị viên có thể cấu hình cho phép hệ thống thu thập thông các trang web khác

### 1.1.2 Yêu cầu đặt ra

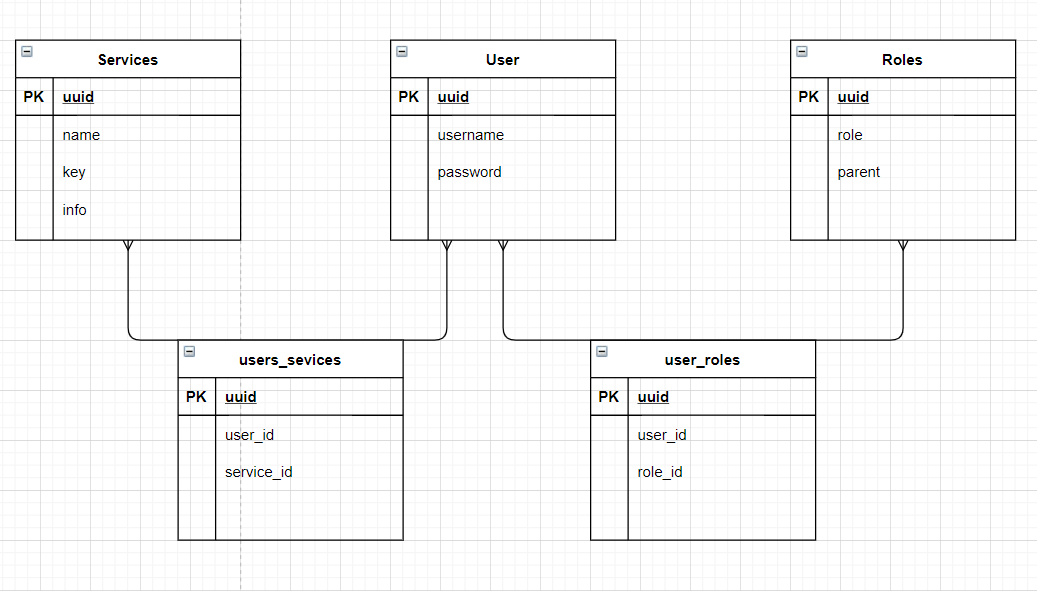
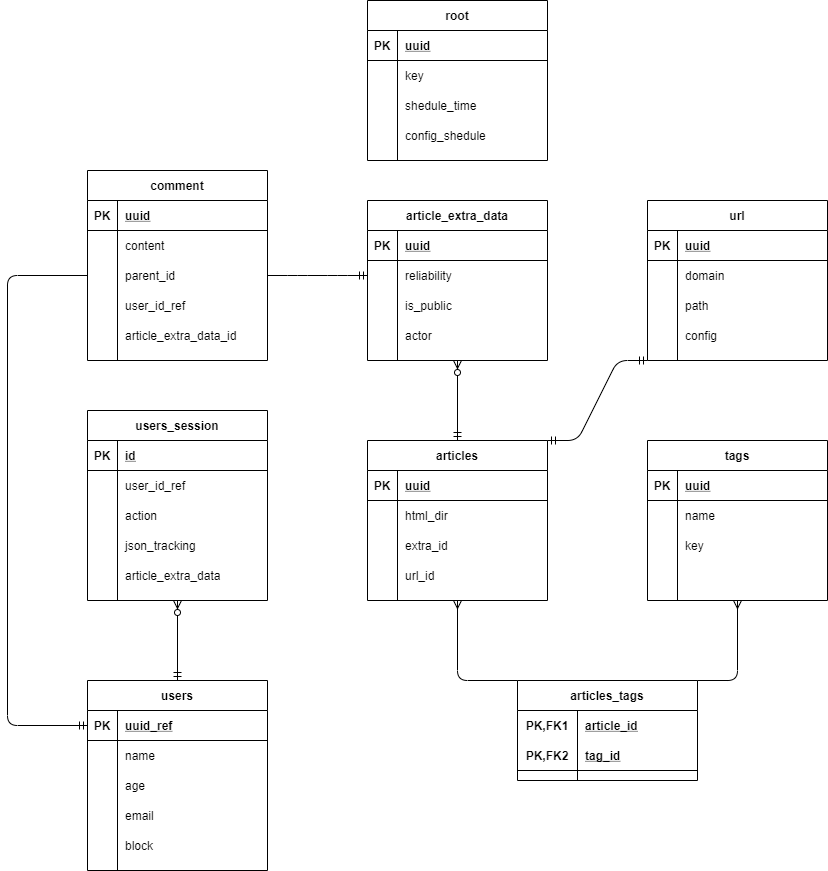
* Hệ thống sẽ gồm 3 phần

1. Phần hiển thị tin tức trực tuyến:
   1. Hiện thị các bài báo dựa vào sở thích đọc tin tức của người dùng thông qua việc sử dụng trang web
   2. Chức năng tìm kiếm bài bào dựa vào các tiêu chí
   3. Chức năng bình luận và đánh giá chất lượng bài báo
2. Phần Thu thập thông tin
   1. Quản trị viên sẽ thiết kế được việc thu thập thông tin tại các trang web nào
   2. Quan trị viên sẽ theo dõi đươc qua trình thu thập thông tin
   3. Hệ thống sẽ trải qua các bước để thu thập thông tin: tải mã nguồn, phân tích bóc tách mã nguồn, lưu trữ mã nguồn
3. Phân xác thực
   1. Thực hiện xác thực để kết hợp với các dịch vụ khác. Nhằm tạo nên 1 hệ sinh thái an toàn

## II. PHÂN TÍCH HỆ THỐNG

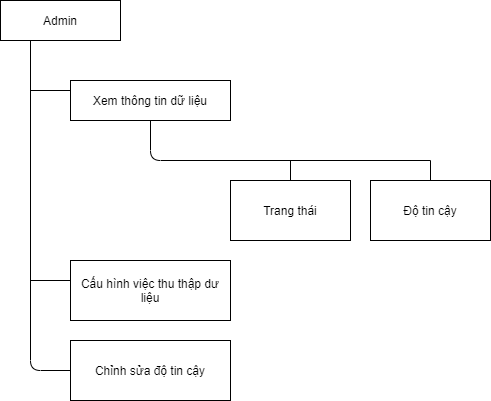
### 1.2.1 Biểu đồ tổng quát hệ thống

### 1.2.2 Cơ sở dữ liệu

1. Cơ sở dữ liệu của AuthenService
2. Cơ sở dữ liệu của User và thông tin thu thập

### 1.2.3 Hệ thống thu thập thông tin

1. Sơ đồ phân cấp chức năng



Chức năng: Xem thông tin dữ liệu

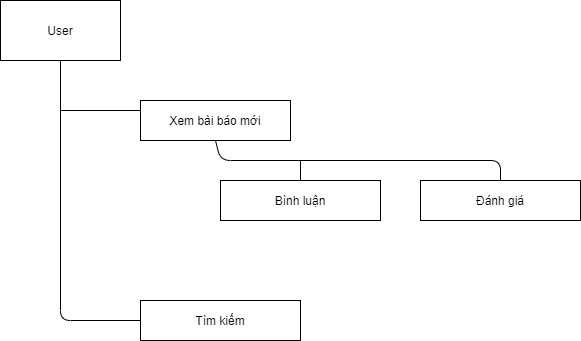
* Dữ liệu sẽ tự động hóa hằng ngày để cập nhập tin tức mới nhất dựa vào các cấu hình có sẵn. Hệ thống sẽ tự đánh giá độ tin cây theo 2 tiêu chí: dựa vào các trang web uy tín do admin lựa chọn ( dantri.com, 24h.com) và dựa vào nội dung của bài báo
* Đầu vào: Ngày tháng
* Các quá trình xử lý
  + Hệ thống sẽ thu thập các thông tin về bài viết mới ở trong cơ sở dữ liệu và hiển thị
* Đầu ra: các bài viết dựa vào ngày tháng.

Chức năng: Cấu hình thu thập dữ liệu

* Bộ selector sẽ đc quản trị viên setup nhằm phục vụ cho việc thu thập dữ liệu.
* Đầu vào: Selector, url
* Các quá trình xử lý
  + Hệ thống sẽ truyền lệnh tới module crawling và đưa thông tin về cấu hình, địa chỉ truy cập
  + Module dựa vào các thông tin nhận được tiến hành thu thập và phân tích dữ liệu sau đó lưu vào trong cơ sở dữ liệu theo ngày, hoặc theo giờ.
  + Các Module crawling chia thành nhiều module nhỏ và hoạt động độc lập ko liên quan tới nhau. Sẽ không ảnh hưởng gì tới người dùng.
* Đầu ra: Thông tin đã qua xử lý trong cơ sở dữ liệu ( nội dung, tiêu đề, loại, độ tin cậy)

### 1.2.4 Hệ thống hiển thị dữ liệu cho người dùng

1. Sơ đồ phân cấp chức năng



Chức năng: Xem thông tin dữ liệu

* Hệ thống sẽ hiển thị menu cho người dùng lựa chọn và đồng thời hiện chi tiết từng bài báo
* Đầu vào:
* Các quá trình xử lý
  + Hệ thống sẽ hiển thị thông tin bài bào dựa vào thông tin người dùng
* Đầu ra: các bài viết

Chức năng: Tìm kiếm

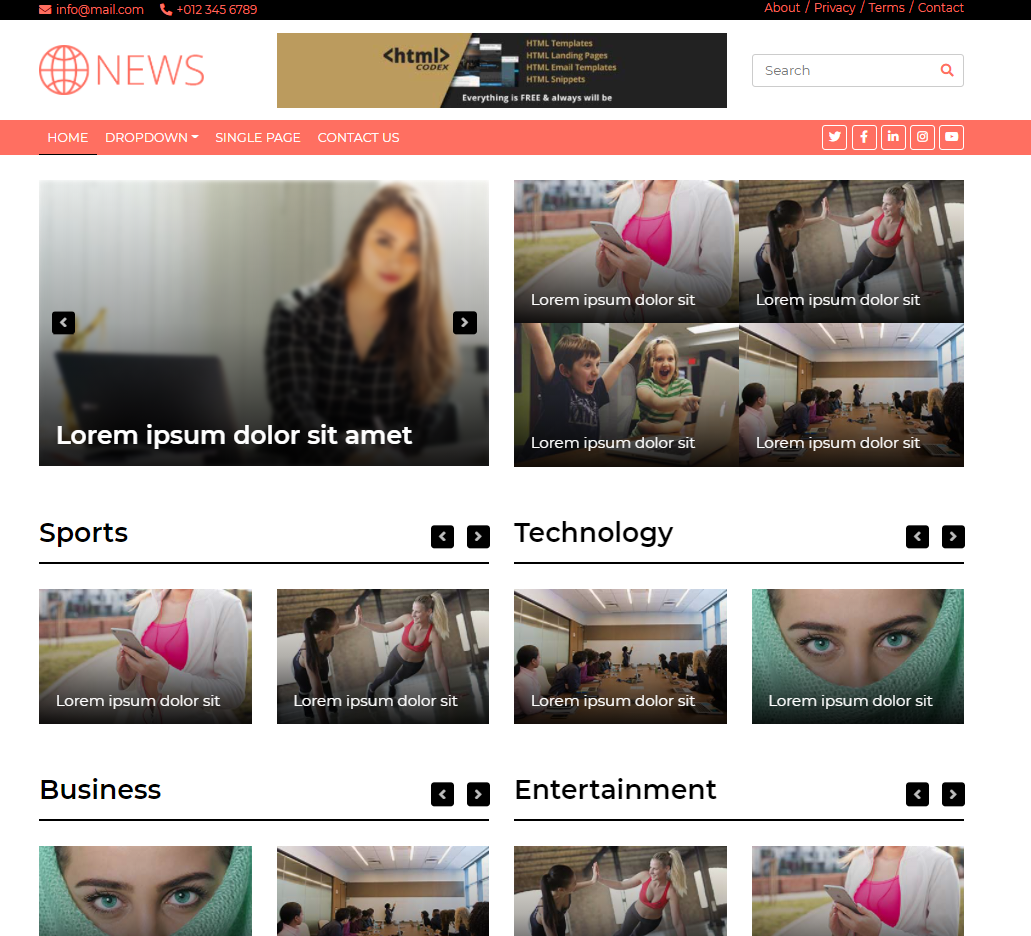
* Dựa vào từ khóa người dùng cung cấp sẽ tìm kiếm trong nội dung bài báo hoặc tiêu đề. Sử dụng cơ sở dữ liệu elasticsearch 1 cơ sở dữ liệu dùng để truy vấn dữ liệu realtime nhanh chóng
* Đầu vào: từ khóa
* Các quá trình xử lý
  + Bóc tách từ khóa và thực hiện multiple search trong tiêu đề và nội dung
* Đầu ra: Các bài báo mà có thể liên quan tới từ khóa

## II. Thực hiện và triển khai

### Giao diện triển khai

Giao diên về người dùng sử dụng để đọc thông tin

* 1. Giao diện trang chủ



* 1. Giao diện bài báo chi tiết

