



Mùa hè sáng tạo 2012

TỰ ĐỘNG HÓA BÓC VÀ LẤY TIN CHO TRANG TIN ĐIỆN TỬ TỔNG HỢP

NỘI DUNG TRÌNH BÀY



- 1 Đặt vấn đề
- 2 Hướng tiếp cận
- 3 Nội dung nghiên cứu
- 4 Kết luận
- 5 Hướng phát triển
- 6 Demo sản phẩm

1. ĐẶT VẤN ĐỀ



- Số lượng các trang báo điện tử ngày càng tăng.
- Người sử dụng truy cập để tìm kiếm thông tin.
- Để có nội dung phong phú, cập nhật thông tin liên tục là một việc không đơn giản.

1. ĐẶT VẤN ĐỀ



1. ĐẶT VẤN ĐỀ



Công việc này hoàn toàn có thể tự động hóa bằng Module GetNews chạy trên nền NukeViet.

2. HƯỚNG TIẾP CẬN



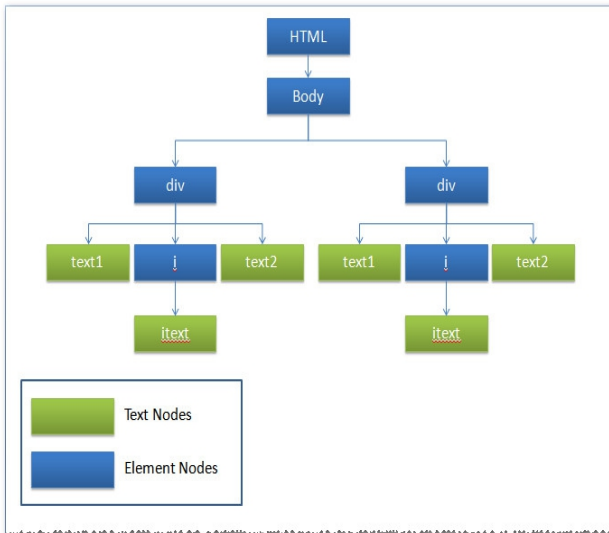
Là hệ quản trị nội dung nguồn mở do người Việt phát triển, NukeViet CMS cho phép xây dựng và quản lý các website, cổng thông tin điện tử rất dễ dàng

2. HƯỚNG TIẾP CẬN



RSS là một tiêu chuẩn định dạng tài liệu dựa trên XML tóm lược thông tin vào một đoạn dữ liệu ngắn gọn, hợp chuẩn.

2. HƯỚNG TIẾP CẬN



DOM Document là một tập của các nút, hay những mẫu thông tin, được tổ chức trong một hệ phân cấp dạng cây.

3. CHỨC NĂNG CHÍNH



- Thêm nguồn tin: Thêm thông số cho bộ lõi lấy tin:
 - Tự động: Hệ thống tự động tìm ra đường dẫn tới tiêu đề, mô tả, nội dung.
 - Thủ công: Người sử dụng tự chọn các thành phần (tiêu đề, mô tả, nội dung), sau đó hệ thống tự tìm ra đường dẫn.

3. CHỨC NĂNG CHÍNH



- Thêm nguồn tin: Thêm thông số cho bộ lõi lấy tin:
 - Tự động: Hệ thống tự động tìm ra đường dẫn tới tiêu đề, mô tả, nội dung.
 - Thủ công: Người sử dụng tự chọn các thành phần (tiêu đề, mô tả, nội dung), sau đó hệ thống tự tìm ra đường dẫn.

3. CHỨC NĂNG CHÍNH



- Lấy tin: Theo 2 phương pháp: Tự động và thủ công.
 - Tự động: Sau một khoảng thời gian, tự động truy cập tới trang nguồn, lọc ra những tin mới.
 - Thủ công: Người sử dụng chọn những tin muốn lấy từ những tin đăng trên trang nguồn.

3. CHỨC NĂNG CHÍNH



- Quản trị tin tức đã lấy: Người sử dụng có thể lựa chọn đăng tin hoặc xóa những tin mà hệ thống đã lấy được.

3. CHỨC NĂNG CHÍNH



- Thống kê lấy tin: Đưa ra những biểu đồ thống kê số lượng tin đã lấy được.
 - Thông qua thống kê đó, người sử dụng có thể thay đổi tần suất lấy tin tự động hoặc chuyên mục lưu.

4. KẾT LUẬN:



- Hiểu biết hơn về hệ thống nukeviet CMS và được học và làm việc với github.
- Module bổ sung thêm tính năng cho mã nguồn nukeviet
- Module có tính ứng dụng cao giúp nâng cao hiệu quả lao động, giúp tiết kiệm thời gian và nguồn nhân lực.

4. KẾT LUẬN:



- Module đã được đưa vào áp dụng trong thực tế và nhận được nhiều sự quan tâm của người dùng(gần 2000 lượt tải trên kho ứng dụng của nukeviet).
- Có tài liệu hướng dẫn sử dụng thuận tiện.

5. HƯỚNG PHÁT TRIỂN:



- Tối ưu bộ lọc lấy tin để mở rộng số lượng trang tin có thể lấy.
- Phát triển robot tìm kiếm, phân lớp nội dung dựa trên phân tích ngữ nghĩa.

6. DEMO SẢN PHẨM:



XIN CHÂN THÀNH CẢM ƠN!