

KỸ THUẬT THU THẬP DỮ LIỆU TỰ ĐỘNG TRÊN INTERNET

TS.Phạm Việt Bình, ThS.Nguyễn Văn Huân, KS.Đỗ Văn Đại

TÓM TẮT:

Trong những năm gần đây Internet đã trở thành một trong những kênh về khoa học, thông tin kinh tế, thương mại và quảng cáo. Thông tin trên các trang Website đa dạng về mặt nội dung cũng như hình thức, có thể nói Internet như một xã hội ảo, nó bao gồm các thông tin về mọi mặt của đời sống kinh tế, xã hội được trình bày dưới dạng văn bản, âm thanh, hình ảnh..Tuy nhiên cùng với ự đa dạng và số lượng lớn thông tin như vậy đã nảy sinh vấn đề quá tải thông tin. Do đó đặt ra một vấn đề là làm thế nào để có thể tìm kiếm, phân loại, xử lý từ nguồn dữ liệu văn bản đó. Mặt khác với dữ liệu ta thu thập, lưu trữ và xử lý được ta có thể phân tích số liệu về thương mại, chứng khoán của từng giai đoạn khác nhau. Trên cơ sở này tôi nghiên cứu và xây dựng hệ thống thu thập, đánh giá và phân cụm thông tin tự động trên Internet để thu thập, lưu trữ, xử lý và phân phối thông tin giúp cho người đọc có thể nắm bắt được thông tin một cách tổng quát.