**LỜI MỞ ĐẦU**

Trong khóa luận này, chúng tôi phát triển một hệ thống dùng để cập nhật kho dữ liệu chỉ mục bài báo khoa học trong lĩnh vực khoa học máy tính bằng cách kết hợp giữa thu thập rút trích thông tin chỉ mục các bài báo trực tiếp từ thư viện số để bổ sung vào dữ liệu chỉ mục có sẵn, hướng tới có được một kho dữ liệu thông tin bài báo đầy đủ, chính xác và cập nhật.

Bằng cách rút trích, thu thập thông tin chỉ mục các bài báo trực tiếp trên các thư viện số, hệ thống đảm bảo thông tin thu thập được là chính xác và cập nhật. Cùng với đó, hệ thống sẽ sử dụng các cơ sở dữ liệu chỉ mục có sẵn để dữ liệu xây dựng được là đầy đủ.

Hệ thống trong khóa luận thực hiện rút trích thông tin chỉ mục trên các thư viện số ACM, IEEEXplore, CiteSeer và sử dụng cơ sở dữ liệu chỉ mục có sẵn DBLP. Cùng với chức năng tự động cập nhật các bài báo mới từ các thư viện số đảm bảo thông tin chỉ mục các bài báo khoa học thu thập được sẽ chính xác, đầy đủ và cập nhật.