1. Khảo sát các công cụ liên quan :

Với mục đích tổ chức lại dữ liệu nhằm hỗ trợ, giúp đỡ cho việc cho việc tìm kiếm nhanh các tài liệu dựa trên những trường metadata.

Greenstone[[1]](#footnote-1) là một bộ phần mềm dùng để xây dựng và phân phối các bộ sưu tập thư viện số. Nó cung cấp một phương pháp mới để tổ chức và xuất bản thông tin trên Internet hoặc trên CD-ROM. Greenstone là sản phẩm của dự án **New Zealand Digital Library** của trường đại học **University of Waikato**, được phát triển và phân phối với sự tham gia của **UNESCO** và **Human Info NGO**. Mục tiêu của phần mềm là cung cấp cho người sử dụng, đặc biệt là các trường đại học, các thư viện hoặc trong các trụ sở công cộng xây dựng các thư viện số riêng của họ.

JabRef[[2]](#footnote-2) là một phần mềm quản lý tài liệu tham khảo sử dụng BibTeX như định dạng gốc của nó. JabRef cung cấp một giao diện dễ sử dụng cho việc chỉnh sửa các tập tin BibTeX, để nhập dữ liệu từ cơ sở dữ liệu khoa học trực tuyến như : CiteSeer, ACM portal, ArXiv, … , và để quản lý và tìm kiếm các tập tin BibTeX[[3]](#footnote-3).

Những công cụ trên cũng cùng chức năng là tổ chức dữ liệu còn có rất nhiều công dụ khác (tham khảo : <http://en.wikipedia.org/wiki/Comparison_of_reference_management_software> ).

Các công cụ được đề cập chỉ mang tính tổ chức dữ liệu metadata đã có sẵn và chuyển thành cơ sở dữ liệu mang tính quãn lý thông tin hơn. Metadata Extraction Tool[[4]](#footnote-4) là một công cụ tạo ra, hình thành metadata cho một tập tin dữ liệu, từ đó mới đi đến việc tổ chức các metadata này thành cơ sở dữ liệu để quản lý. Nhưng Metadata Extractio tool chỉ dừng lại ở việc lấy những thông tin thuộc tính của tập tin, ví dụ : định dạng (.doc, pdf, xml,…), kích thước tập tin (bao nhiêu KB, …), tên tập tin, …và hình thành tổ hợp thông tin cho tập tin đó, biểu diễn thành XML.

Với mong muốn cao hơn công cụ của nhóm thực hiện sẽ đi vào nội dung bên trong của tập tin, tìm kiếm, nhận dạng thông minh và chính xác các thông tin metadata của bài báo khoa học. Từ đó hình thành tập tin XML để hổ trợ cho việc quãn lý metadata sau này.

1. http://www.greenstone.org/ [↑](#footnote-ref-1)
2. *http://jabref.sourceforge.net/* [↑](#footnote-ref-2)
3. BibTeX was created by [Oren Patashnik](http://en.wikipedia.org/wiki/Oren_Patashnik) and [Leslie Lamport](http://en.wikipedia.org/wiki/Leslie_Lamport) in 1985. [↑](#footnote-ref-3)
4. *http://meta-extractor.sourceforge.net/* [↑](#footnote-ref-4)