

GIỚI THIỆU CÔNG CỤ HỖ TRỢ NGHIÊN CỨU KHOA HỌC: CÁC PHẦN MỀM R VÀ ENDNOTE

Nguyễn Minh Tâm¹, Hoàng Anh Đào²

1. Khoa Y tế Công cộng, 2. Khoa Răng Hàm Mặt, Trường Đại học Y Dược Huế

Tóm tắt:

Các phần mềm máy tính đã và đang góp phần tích cực vào sự phát triển của khoa học. Bài viết này giới thiệu hai phần mềm đã và đang được đánh giá cao trong việc hỗ trợ phân tích thống kê (phần mềm R) và quản lý tài liệu tham khảo (phần mềm EndNote). Với mỗi phần mềm, bài viết sẽ giới thiệu về các đặc điểm, một số chức năng chính của phần mềm và các tài liệu tham khảo (chủ yếu là các websites miễn phí). Việc sử dụng các phần mềm này sẽ giúp các nhà nghiên cứu ở các nước đang phát triển như Việt Nam tiết kiệm được chi phí (phần mềm R) cũng như nâng cao chất lượng các bài báo khoa học (phần mềm EndNote).

Abstract:

AN INTRODUCTION TO TOOLS USED IN RESEARCH: R AND ENDNOTE SOFTWARES

Nguyen Minh Tam, Hoang Anh Dao

Computer softwares have made great contributions to the development of science. This paper introduces two softwares that have been used widely in statistics (R software) and in reference management (EndNote software). For each software, there will be a description of main functions and a list of further readings (free websites). Using these two software might help researchers in developing countries like Vietnam to save budget (R software) as well as to improve the quality of their paper (EndNote software).

1. Phần mềm thống kê R

Việc ứng dụng máy tính cho phân tích và tính toán đã giúp khoa học thống kê có một bước phát triển nhảy vọt và góp phần tích cực vào sự phát triển của khoa học thực nghiệm. Trong thời gian qua, việc phân tích số liệu và vẽ biểu đồ thường được thực hiện bằng các phần mềm thông dụng như *SAS*, *SPSS*, *Stata*, *Statistica*, và *MedCalc*. Tuy nhiên, có một trở ngại cho người sử dụng là chi phí để sử dụng các phần mềm này, khiến ngay cả các trường Đại học và trung tâm nghiên cứu của các nước phát triển cũng khó khăn trong việc duy trì mua bản quyền lâu dài. Vấn đề này thực sự là một trở ngại lớn cho các trường đại học và trung tâm nghiên cứu ở các nước đang phát triển và giải pháp phổ biến là sử dụng các bản

bẻ khóa (crack) của các phần mềm này. Điều đáng mừng là tình trạng này đã và đang được khắc phục và có thể chấm dứt hoàn toàn trong tương lai khi có sự xuất hiện của R - một phần mềm miễn phí được đánh giá là đã, đang và sẽ làm một cuộc cách mạng trong ngành thống kê học. Trong phạm vi bài viết này, xin giới thiệu tóm tắt về R và một số nguồn tài liệu tham khảo.

- Quá trình phát triển:

R được bắt đầu viết vào năm 1996 bởi Robert Gentleman và Ross Ihaka lúc bấy giờ thuộc Bộ môn Thống kê Trường Đại học Auckland, New Zealand. Sau đó, R đã được các nhà thống kê trên thế giới hợp lực phát triển với chủ trương mã nguồn mở, sao cho tất cả các thành viên trong ngành thống kê học và

toán học trên thế giới có thể sử dụng một cách thông nhất và **hoàn toàn miễn phí**. Cho đến nay, qua gần 15 năm phát triển, ngày càng có nhiều nhà thống kê học, toán học, nghiên cứu trong mọi lĩnh vực đã chuyển sang sử dụng R để phân tích dữ liệu khoa học. Trên toàn cầu, đã có một mạng lưới hơn một triệu người sử dụng R, và con số này đang tăng rất nhanh. R có thể tải xuống máy tính cá nhân của bất cứ cá nhân nào, bất cứ lúc nào và bất cứ ở đâu trên thế giới. Chỉ vài phút cài đặt là R có thể đưa vào sử dụng. Chính vì thế mà đa số các đại học trên thế giới đang chuyển sang sử dụng R cho học tập, nghiên cứu và giảng dạy. Có nhiều nhận xét cho rằng với sự phát triển của R, trong vòng 10 năm nữa, vai trò của các phần mềm thống kê thương mại sẽ không còn lớn như trong thời gian qua.

- R là gì?

Nói ngắn gọn, R là một phần mềm sử dụng cho xử lý số liệu, phân tích thống kê và vẽ biểu đồ. Nói rộng hơn, về bản chất, R là ngôn ngữ máy tính đa năng, có thể sử dụng cho nhiều mục tiêu khác nhau, từ tính toán đơn giản, toán học giải trí (recreational mathematics), tính toán ma trận (matrix), cho đến các phân tích thống kê phức tạp. Vì là một ngôn ngữ, cho nên người ta có thể sử dụng R để phát triển thành các phần mềm chuyên môn cho một vấn đề tính toán cá biệt. Tuy là miễn phí nhưng chất lượng của R rất cao, mà cụ thể là R có khả năng làm tất cả những phân tích

mà các phần mềm thương mại làm được.

- Tải R xuống và cài đặt vào máy tính

Để cài đặt R, truy cập website có tên là “Comprehensive R Archive Network” (CRAN) <http://cran.R-project.org> và tải về máy tính. Phiên bản mới nhất của R là **R version 2.12.1** được giới thiệu vào ngày 16 tháng 12 năm 2010. Cũng tại website này, có thể tìm thấy rất nhiều tài liệu chỉ dẫn cách sử dụng R, đủ trình độ, từ nhập môn cho đến cao cấp.

- Tải R Package và cài đặt vào máy tính

R cung cấp cho người sử dụng một “ngôn ngữ” máy tính và một số hàm (*function*) để làm các phân tích căn bản và đơn giản. Ưu điểm của R đó là người sử dụng có thể phát triển các ‘package’, tức là các phần mềm nhỏ được phát triển để giải quyết một vấn đề cụ thể, và có thể chạy trong hệ thống R. Người sử dụng cũng có thể tải về máy tính các *package* đã được phát triển bởi người khác (thường là các nhà thống kê học) để làm những phân tích phức tạp hơn. Ví dụ, để phân tích hồi qui tuyến tính, R có *function lm* để sử dụng cho mục đích này, nhưng để làm các phân tích sâu hơn và phức tạp hơn, chúng ta cần đến các package như **lme4**. Các package này cần phải được tải về và cài đặt vào máy tính. Để tải các package, đăng nhập vào website <http://cran.r-project.org>, chọn phần “**Packages**” ở bên trái của mục lục trang web. Theo GS. Nguyễn Văn Tuấn, một số package cần tải về máy tính để sử dụng cho các phân tích dịch tễ học là:

Tên Package	Chức năng
<i>Trellis</i>	Dùng để vẽ đồ thị và làm cho đồ thị đẹp hơn
<i>Lattice</i>	Dùng để vẽ đồ thị và làm cho đồ thị đẹp hơn
<i>Hmisc</i>	Một số phương pháp mô hình dữ liệu của F. Harrell
<i>Design</i>	Một số mô hình thiết kế nghiên cứu của F. Harrell
<i>Epi</i>	Dùng cho các phân tích dịch tễ học
<i>Epi tools</i>	Một số Package chuyên cho các phân tích dịch tễ học
<i>Foreign</i>	Dùng để nhập dữ liệu từ các phần mềm khác như SPSS, Stata, SAS, v.v...
<i>Rmeta</i>	Dùng cho các phân tích tổng hợp (meta-analysis)
<i>Meta</i>	Một Package khác cho phân tích tổng hợp

<i>Survival</i>	Chuyên dùng cho phân tích theo mô hình Cox (Cox's proportional hazard model)
<i>Zelig</i>	Package dùng cho các phân tích thống kê trong lĩnh vực xã hội học
<i>Genetics</i>	Package dùng cho phân tích số liệu di truyền học
<i>BMA</i>	Bayesian Model Average

Các package này có thể cài đặt trực tuyến bằng cách chọn **Install packages** trong phần **packages** của R, hoặc nếu package đã được tải xuống máy tính cá nhân thì việc cài đặt có thể nhanh hơn bằng cách chọn **Install package (s) from local zip file** cũng trong phần **packages**.

- Cách nhập dữ liệu vào R

Dữ liệu dùng cho R là dữ liệu dạng data.frame. Có nhiều cách để nhập số liệu vào một data.frame trong R, từ nhập trực tiếp đến chuyển đổi từ các nguồn khác. Một ưu điểm của R là tính tương tác với bộ số liệu của các phần mềm thống kê khác như Excel, SPSS, Stata, SAS. Một số cách thông dụng là:

- Nhập số liệu trực tiếp
- Nhập số liệu từ một file dạng *text* bằng lệnh *read.table*
- Nhập số liệu từ các phần mềm khác với sự hỗ trợ của package Foreign, như nhập từ Excel bằng lệnh *read.csv*, từ một file của SPSS bằng lệnh *read.spss*,

- Sử dụng R trong phân tích số liệu và vẽ biểu đồ

R được sử dụng vào các mục đích và hoàn cảnh khác nhau rất đa dạng, từ việc sử dụng như một máy tính cầm tay cho đến việc sử dụng cho các phép tính ma trận và lập chương. R có thể sử dụng để làm sạch số liệu cũng như tập trung số liệu thành một nhóm, hay tách rời thành từng nhóm, hay thay thế từ kí tự (characters) sang số (numeric) cho tiện việc tính toán. Đối với Phân tích thống kê, R có thể sử dụng cho Thống kê mô tả (descriptive statistics, summary); thực hiện các Kiểm định như t (t.test), kiểm định Wilcoxon; Chi bình phương (Chi squared test, chisq.test); hay

Kiểm định Fisher (Fisher's exact test, fisher.test), hay dùng để thực hiện So sánh nhiều tỉ lệ (prop.test, chisq.test). R thực hiện rất tốt các phân tích phức tạp như Phân tích hồi qui tuyến tính (linear regression), Phân tích phương sai (analysis of variance - ANOVA), Phân tích hồi qui logistic và có thể sử dụng trong ước tính cỡ mẫu (sample size estimation). Đặc biệt, trong ngôn ngữ R có rất nhiều cách để thiết kế một biểu đồ gọn và đẹp. Phần lớn những hàm để thiết kế biểu đồ có sẵn trong R, còn với số loại biểu đồ tinh vi và phức tạp khác thì có thể thiết kế bằng các package chuyên dụng như lattice hay trellis có thể tải từ website của R.

- Tài liệu tham khảo về R

Hiện nay, thư viện sách về R còn tương đối khiêm tốn so với thư viện cho các phần mềm thương mại như SAS và SPSS. Tuy nhiên, các chỉ dẫn về cách sử dụng R hiện có trên website của R-project cũng như từ các trường đại học và website cá nhân trên khắp thế giới cũng đã cung cấp một lượng thông tin phong phú và có thể thỏa mãn nhu cầu tự học của đa số người sử dụng. Dưới đây là một số sách và nguồn tham khảo về R.

- Ihaka R, Gentleman R. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996; 5:299-314: Bài báo của hai người sáng tạo ra R.

- Website chính thức của R: <http://cran.R-project.org>, trong đó có một số tài liệu quan trọng như “**An Introduction to R**” của W. N. Venables và B. D. Ripley (<http://cran.r-project.org/doc/manuals/R-intro.pdf>.)

- “**Phân tích số liệu và tạo biểu đồ bằng**

R: hướng dẫn và thực hành”, sách của Nguyễn Văn Tuấn do Nhà xuất bản Đại học Quốc gia Thành phố Hồ Chí Minh xuất bản năm 2006. Phần tóm lược một số chương chính của cuốn sách này có thể tìm thấy ở website: www.R.ykhoa.net. Trang web này còn có tất cả các dữ liệu (datasets) và các mã sử dụng trong sách để bạn đọc có thể tải xuống máy tính cá nhân để sử dụng.

“Data Analysis and Graphics Using R – An Example Approach” (Nhà xuất bản Cambridge University Press, 2003) của John Maindonald nay đã tái bản lần thứ 2 với thêm một tác giả mới John Braun. Đây là cuốn sách rất có ích cho những ai muốn tìm hiểu và học về R. Năm chương đầu của sách viết cho bạn đọc chưa từng biết về R, còn các chương sau thì viết cho các bạn đọc đã biết cách sử dụng R thành thạo.

“Introductory Statistics With R” (Nhà xuất bản Springer, 2004) của Peter Dalgaard. Đây là cuốn sách loại căn bản cho R, chỉ khoảng 200 trang nhưng rất bổ ích.

“Linear Models with R” (Nhà xuất bản Chapman & Hall/CRC, 2004) của Julian Faraway, 213 trang. Sách hiện có thể tải từ internet xuống miễn phí tại website sau đây:

<http://www.stat.lsa.umich.edu/~faraway/book/pr.pdf>

hay <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>.

“R Graphics (Computer Science and Data Analysis)” (Nhà xuất bản Chapman & Hall/CRC, 2005) của Paul Murrell. Đây là cuốn sách chuyên về phân tích biểu đồ bằng R, bao gồm rất nhiều mã để bạn đọc có thể tự mình thiết kế các biểu đồ phức tạp.

Một số tài liệu hướng dẫn cách sử dụng R có thể tải miễn phí và tham khảo:

- **“R for Beginners”** (57 trang) của Emmanuel Paradis. Tài liệu được soạn cho bạn đọc mới làm quen với R.

Địa chỉ internet: http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf.

- **“Using R for Data Analysis and Graphics: Introduction, Code and Commentary”** (35 trang) của John Maindonald là một tóm lược các lệnh và hàm căn bản của R cho phân tích số liệu và biểu đồ.

Địa chỉ internet: <http://cran.r-project.org/doc/contrib/usingR.pdf>

- **“Statistical Analysis with R – a quick start”** (46 trang) của Oleg Nenadic và Walter Zucchini. Địa chỉ internet:

http://www.stat.oek.wiso.unigoettingen.de/mitarbeiter/ogi/pub/r_workshop.pdf

- **“Notes on the use of R for psychology experiments and questionnaires”** (39 trang) của Jonathan Baron và Yuelin Li. Web. Tài liệu được soạn cho giới nghiên cứu tâm lý học và xã hội học. Có ví dụ về log - linear model và một số mô hình phân tích phương sai trong tâm lý học.

Địa chỉ internet: <http://www.psych.upenn.edu/~baron/rpsych/rpsych.html>

- **StatsRus** gồm một sưu tập về các mẹo để sử dụng R hữu hiệu hơn, dài 80 trang. Địa chỉ internet: <http://lark.cc.ukans.edu/pauljohn/R/statsRus.html>

2. Phần mềm EndNote

EndNote là chương trình Quản lý Tài liệu tham khảo phổ biến nhất hiện nay. Các chức năng chính của chương trình này là:

- Quản lý một cơ sở dữ liệu cá nhân về các tài liệu tham khảo.

- Tải (download) các tài liệu tham khảo từ các cơ sở dữ liệu khác.

- Sử dụng cơ sở dữ liệu cá nhân (về tài liệu tham khảo) để chèn (insert) các tài liệu tham khảo vào các văn bản dạng word.

- Tạo mục Tài liệu tham khảo hoặc bibliography theo đúng thể loại (style) để xuất bản.

“Thư viện EndNote” (EndNote Library):

- “Thư viện EndNote” là một file cho

phép lưu và quản lý một cơ sở dữ liệu các tài liệu tham khảo. Một thư viện có thể chứa tới 100.000 tài liệu tham khảo. Một thư viện cũng là một file như các file khác trong máy tính. Nó có thể được copy, thay đổi tên, xóa, di chuyển sang thư mục khác hay nén lại (compressed).

- Có thể lập và duy trì nhiều thư viện nhưng điều này là không cần thiết, mỗi người sử dụng chỉ cần lập một thư viện là đủ. Tất cả các tài liệu tham khảo có thể lưu giữ trong một thư viện, và được dán nhãn (labeled) hoặc nhóm (grouped) để phân biệt các chủ đề, nội dung khác nhau.

- Trong mỗi thư viện EndNote, người dùng có thể biên tập, nhận xét, ghi chú cho mỗi tài liệu tham khảo cũng như ý định sử dụng các tài liệu này sau này. Đối với mỗi tài liệu, bên cạnh lưu trữ các thông tin hành chính và tóm tắt (abstract), người dùng có thể dán (paste) file toàn văn của tài liệu dưới dạng .pdf vào Thư viện EndNote. Điều này rất tiện lợi cho người sử dụng trong việc xem lại các tài liệu, đặc biệt là vẫn sử dụng được ngay cả khi không truy cập được internet (offline).

- **Lưu ý:** Thư viện EndNote có thể copy và lưu giữ trong máy tính hay USB hoặc chia sẻ qua email. Tuy nhiên, khi copy cần lưu ý copy cả thư mục dữ liệu .rdb. Hơn thế nữa, khi chia sẻ cần lưu ý là version đang sử dụng phải tương thích với nhau. Các version gần đây không tương thích với EndNote 7 và các versions trước đó.

Một số chức năng chính thường được sử dụng trong EndNote

- **Nhập tài liệu tham khảo vào Thư viện EndNote**

Các tài liệu tham khảo có thể được nhập trực tiếp (một cách thủ công) bằng cách gõ

vào các phần tương ứng. Cách này thường áp dụng cho các trường hợp tài liệu tham khảo là sách, tạp chí in và không có trên internet. Đối với các tài liệu tham khảo có ở các cơ sở dữ liệu trên web, có thể chọn và chuyển trực tiếp (direct export) vào Thư viện EndNote. Cách này nhanh hơn nhiều so với cách nhập thủ công, tuy nhiên chỉ thực hiện được khi dùng trình duyệt **Internet Explorer** hoặc **Firefox**. Với một số cơ sở dữ liệu, như Web of Science, hình thức nhập này thực hiện rất tốt. Tuy nhiên, có một số cơ sở dữ liệu thì cần phải chỉnh sửa lại sau khi nhập.

Ví dụ về nhập các tài liệu tham khảo từ Web of Science: sau khi đăng nhập vào Web of Science, gõ tìm các tài liệu theo các từ khóa, trang web sẽ hiển thị danh mục các tài liệu tham khảo tìm được. Đánh dấu (mark) vào hộp nhỏ bên cạnh các tài liệu tham khảo muốn chọn. Nhấn nút **Save to EndNote, RefMan, ProCite**. Từ đây sẽ kích hoạt quá trình nhập các tài liệu tham khảo từ cơ sở dữ liệu vào Thư viện EndNote. Sau khi hoàn thành, trên Thư viện EndNote sẽ thấy các tài liệu tham khảo vừa được nhập. Các tài liệu này được lưu ở nhóm tạm thời gọi là **Imported References** group. Để hiển thị tất cả các tài liệu tham khảo trong thư viện, nhấn nút **All References** ở thanh công cụ của Thư viện. Lưu ý là phải nhập các tài liệu tham khảo được chọn vào Thư viện EndNote trước khi chuyển sang trang hiển thị kết quả tiếp theo.

- **Tạo một Bibliography hay danh mục tài liệu tham khảo**

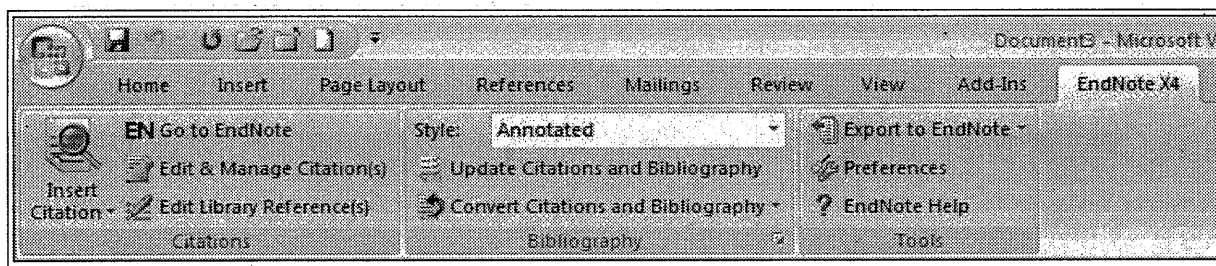
Phần mềm EndNote có khả năng chuyển các dữ liệu lưu trong Thư viện EndNote thành các tài liệu tham khảo được format theo quy định về danh mục tài liệu

tham khảo (output styles) do các tạp chí hoặc các trường đại học, viện nghiên cứu quy định. Phần mềm EndNote có sẵn rất nhiều các output styles, đồng thời có thêm hàng trăm styles khác có thể downloaded từ trang web của EndNote, từ đó có thể chọn ra các output styles phù hợp. Nếu không, EndNote cho phép người sử dụng tạo ra các loại output styles riêng hoặc chỉnh sửa (edit) một output style có sẵn.

- Sử dụng EndNote với văn bản Word

Một điểm mạnh của EndNote là khả năng trích xuất và chèn tài liệu tham khảo từ Thư viện EndNote vào các văn bản Word

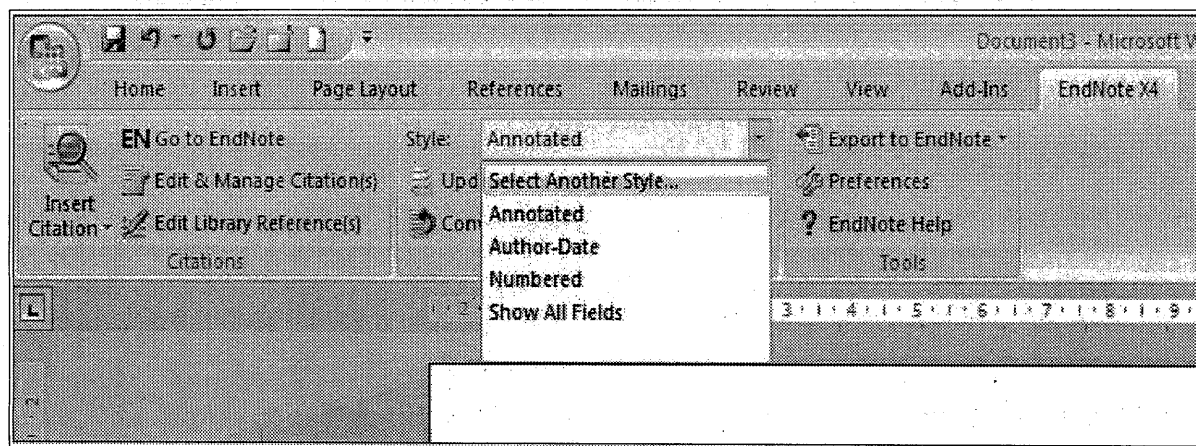
nhờ chức năng **Cite While You Write** (trích dẫn khi đang soạn thảo). Chức năng này cho phép chèn tài liệu tham khảo vào văn bản đồng thời tự động tạo danh mục tài liệu tham khảo ở phần cuối văn bản. Phần mềm EndNote (bản mới nhất là EndNote X4) tương thích với Word 2010, Word 2007 và Word 2003 (hệ điều hành Windows) và Word 2004 và Word 2008 (hệ điều hành Macintosh). Sau khi cài đặt EndNote vào máy tính và kích hoạt ở Word, các chức năng của EndNote sẽ xuất hiện ở thanh công cụ của Word. Hình thức của EndNote sẽ khác nhau ở các version Word khác nhau. Ở Word 2007, khi mở file Word sẽ thấy EndNote xuất hiện như sau



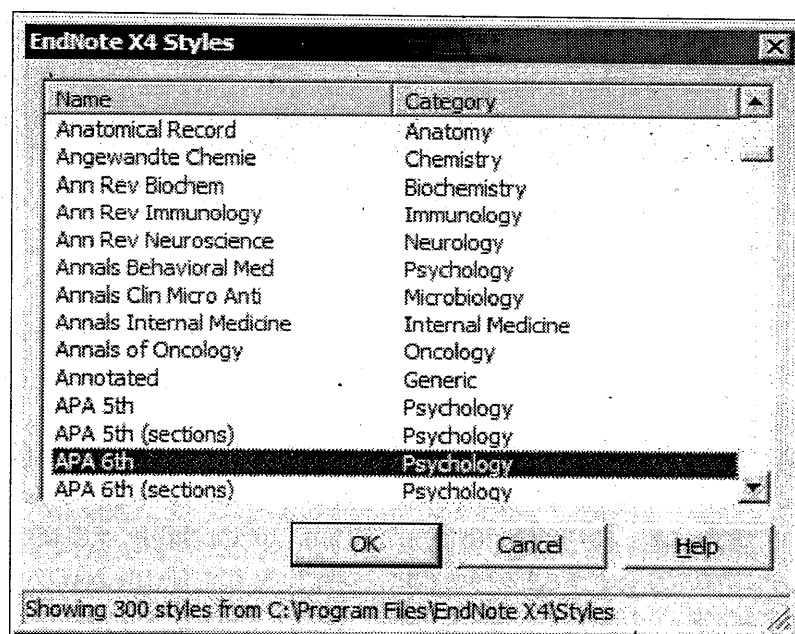
Chức năng này đặc biệt hữu ích khi viết các phần Tổng quan tài liệu vì cho phép chúng ta thực hiện cùng lúc các công việc: tìm các tài liệu liên quan đến phần đang viết, chọn lựa các thông tin cần trích dẫn, chèn các thông tin này vào văn bản và đưa tài liệu vừa được trích dẫn vào danh mục tài liệu tham khảo ở cuối văn bản.

Ví dụ về cách sử dụng EndNote trong khi soạn thảo văn bản Word:

Mở file Word. Đầu tiên cần chọn output style. Ở đây chúng ta thử chọn APA style vốn hay được sử dụng. Trên menu, tìm mục xử lý style và chọn **Select Another Style** (xem hình)



Trên màn hình sẽ hiển thị danh mục các styles để lựa chọn. Nhấn vào phần **Name** để sắp xếp các style theo tên rồi chọn **APA 6th** và nhấn nút **OK**.



Để chèn tài liệu tham khảo vào văn bản, di chuyển đến vị trí cần chèn (thường là cuối câu)



và nhấn nút **Insert Citation** ở menu EndNote trên Word. Một cửa sổ tìm kiếm sẽ hiển thị. Chọn tài liệu tham khảo muốn chèn ở hộp **Search** (thường chọn nhanh hơn bằng cách tìm với tên tác giả). Chọn tài liệu tham khảo muốn chèn và nhấn nút **Insert**. Tài liệu tham khảo sẽ được chèn vào văn bản và EndNote sẽ tự động format theo đúng style đã chọn, ở cả trong văn bản và ở danh mục tài liệu tham khảo cuối văn bản.

Một số địa chỉ Web hữu ích về EndNote

Bài viết này được biên soạn dựa trên tài liệu của các websites dưới đây và xin được giới thiệu để những ai quan tâm có thể tìm hiểu thêm.

- **Trang Web chính thức của EndNote:** <http://www.endnote.com/>

• Cập nhật các thông tin và version mới cho người sử dụng download. Người dùng có

thể tải bản Trial 30 ngày về để sử dụng.

• Trả lời các câu hỏi thường gặp

• Các video clips hướng dẫn sử dụng EndNote bằng hình ảnh (video tutorials)

• Các bộ lọc (filter) và style bổ sung

• Mục tìm kiếm các Style

- **Trang Web hướng dẫn sử dụng EndNote của Thư viện Đại học Queensland:** <http://www.library.uq.edu.au/endnote/>

• Hướng dẫn, giải đáp các thắc mắc thường gặp

• Danh mục các tạp chí và các thuật ngữ thường dùng của các tạp chí chuyên ngành y, sinh, hóa học, và lịch sử.

- **Diễn đàn Thảo luận về EndNote:** <http://forums.thomsonscientific.com/ts/>

• Thảo luận với những người đang sử dụng EndNote

• Tìm kiếm các giải đáp cho vấn đề mình đang gặp mà trước đây diễn đàn đã đăng