
Tools for linguistics research

วิโรจน์ อรุณมานะกุล



5th draft : 2559

ภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทนำ

เอกสารนี้ใช้ประกอบการสอนวิชาเทคนิคการวิจัยทางภาษาศาสตร์ ของภาควิชา
ภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยเน้นที่เรื่องการจัดเก็บ
ข้อมูล และการใช้เครื่องมือหรือ โปรแกรมต่างๆ เพื่อประโยชน์ในการวิจัย
เมื่อกล่าวถึงข้อมูลสำหรับงานวิจัย เราสามารถแยกเป็น ข้อมูลปฐมภูมิ (Primary sources)
กับ ข้อมูลทุติยภูมิ (Secondary sources) ข้อมูลปฐมภูมิ คือข้อมูลดิบที่นักวิจัยรวบรวมมา
เพื่อใช้ในการวิจัย ส่วนข้อมูลทุติยภูมิ คือข้อมูลที่มีผู้เก็บรวบรวมหรือวิเคราะห์มาแล้ว ถ้า
พูดถึงเอกสาร (document) ก็จะเป็น primary sources เป็นเอกสารที่ให้ข้อมูลดิบเพื่อการ
วิจัย ส่วน secondary sources เป็นเอกสารที่ให้ข้อมูลเกี่ยวกับ primary sources อีกทีหนึ่ง
สำหรับข้อมูลในงานวิจัยทางด้านภาษาศาสตร์อาจจะแยกข้อมูลออกเป็นข้อมูล
ภาษา กับ ข้อมูลงานวิจัย ข้อมูลภาษาอาจแยกเป็นข้อมูลประเภทต่างๆ เช่น ข้อมูลภาษา
เขียน ภาษาพูด, ภาษาปัจจุบัน ภาษาเก่า, ภาษามาตรฐาน ภาษาถิ่น, เป็นต้น ข้อมูลภาษา
มีความสำคัญมากเพราะเป็นตัวแทนของปรากฏการณ์ทางภาษาที่เราต้องการศึกษา จึง
ต้องมีปริมาณที่มากพอจะศึกษาได้และเป็นตัวแทนของภาษาที่ต้องการศึกษาอย่างแท้จริง
งานวิจัยจำนวนมากจึงอาศัยการวิเคราะห์ข้อมูลจากคลังข้อมูลภาษาที่เป็นตัวแทนของ
ภาษาที่ต้องการศึกษา

ข้อมูลอาจเก็บจากเอกสารต่างๆ หรือในบางกรณีก็เป็นการเก็บข้อมูลภาคสนาม สิ่ง
ที่ต้องคำนึงคือทำอย่างไรให้เก็บข้อมูลที่ต้องการได้ ไม่มีผลกระทบจากปัจจัยเกี่ยวกับผู้
เก็บข้อมูลและวิธีการเก็บข้อมูล การควบคุมตัวแปรต่างๆ หากมีข้อมูลที่คนอื่นเก็บมาหรือ
สร้างไว้แล้ว เราต้องพิจารณาว่าข้อมูลนั้นตรงกับที่เราต้องการหรือไม่ มีรายละเอียดที่มา

ครบถ้วน และสามารถนำมาใช้ได้โดยตรงหรือไม่ การทำแบบสอบถามผ่านทางอินเทอร์เน็ตก็เป็นวิธีการหนึ่งที่เราสามารถใช้ได้หากต้องการสำรวจข้อมูลต่างๆ เพื่อใช้ในงานวิจัย

นอกจากข้อมูลภาษา ข้อมูลงานวิจัยก็เป็นสิ่งสำคัญ เป็นส่วนของการทบทวนวรรณกรรม ในระหว่างการสำรวจงานต่างๆ ผู้วิจัยจึงต้องเรียนรู้ที่จะเก็บข้อมูลงานวิจัยต่างๆ รู้จักการจัดการข้อมูลบรรณานุกรม การใช้โปรแกรม Endnote เพื่อทำบรรณานุกรม หากไม่มีโปรแกรม Endnote การใช้โปรแกรม Zotero ร่วมกับ browser ต่างๆ ก็เป็นเครื่องมือหนึ่งที่ช่วยในการเก็บระเบียบข้อมูลระหว่างการค้นคว้าได้ดี ในเอกสารนี้จะกล่าวถึงการสร้างรายการบรรณานุกรมใหม่, การโหลดข้อมูลจากห้องสมุด, การเลือกรูปแบบ, การคัดลอกบรรณานุกรมใส่รายงาน โดยใช้โปรแกรม Endnote

การค้นข้อมูลงานวิจัย ปัจจุบันมีฐานข้อมูลจำนวนมากที่มหาวิทยาลัยรับเป็นสมาชิก ในจุฬาฯ หากเข้าไปที่เว็บห้องสมุดจะเห็น CU-reference database ที่เชื่อมโยงไปยังฐานข้อมูลต่างๆ ทำให้สามารถค้น บทความย่อ บทความวิจัย วารสาร หนังสือออนไลน์ นอกจากนี้ ยังมีแหล่งข้อมูลบนอินเทอร์เน็ตอีกเป็นจำนวนมากที่สามารถใช้ประโยชน์ค้นข้อมูลวิจัย เช่น www.scribd.com, google search, scholar.google.co.th,

นอกจากโปรแกรมที่ใช้ในการจัดการข้อมูลต่างๆ แล้ว โปรแกรมสำหรับใช้งานทางภาษาศาสตร์นั้นมีมากมาย ได้แก่ โปรแกรมวาดต้นไม้ โปรแกรมวิเคราะห์เสียง โปรแกรมสำหรับการทดลองทางจิตวิทยา โปรแกรมกำกับข้อมูลภาษา เป็นต้น การทบทวนวรรณกรรมจะช่วยให้เราเห็นว่าเครื่องมืออะไรบ้างที่สามารถใช้ได้ และเครื่องมือไหนเป็นที่ยอมรับใช้กันแพร่หลายในสาขาที่เราสนใจ

นอกจากนี้ สิ่งสำคัญในการวิจัยคือการเก็บรักษาข้อมูลให้ปลอดภัย การสำรองข้อมูล การรักษาความปลอดภัยข้อมูลจึงเป็นสิ่งสำคัญและจำเป็น เพราะเราไม่รู้ว่าพรุ่งนี้จะเปิดคอมพิวเตอร์มาแล้วใช้ได้ไหม จึงต้องมีการสำรองข้อมูลที่ดี หากข้อมูลสูญหายไปด้วยเหตุใด งานวิจัยนั้นก็ล้มเหลวก่อนจะถึงปลายทาง



โปรแกรม Endnote

โปรแกรม Endnote เป็นโปรแกรมสำหรับช่วยทำบรรณานุกรม ข้อดีของโปรแกรมนี้คือสามารถจัดรูปแบบบรรณานุกรมและการอ้างอิงตามที่แต่ละวารสาร หรือแต่ละสมาคมต้องการได้ โดยที่ผู้วิจัยไม่ต้องนั่งพิมพ์หรือจัดรูปแบบใหม่ทุกครั้งที่ต้องการใช้ โปรแกรมยังสามารถคัดลอกข้อมูลบรรณานุกรมจากภายนอก เช่น ห้องสมุด ฐานข้อมูลงานวิจัยต่างๆ หรือจากโปรแกรมอื่นๆ ได้ โดยไม่ต้องเสียเวลาพิมพ์รายการบรรณานุกรมเข้าไปใหม่

การติดตั้ง โปรแกรม EndNote

การติดตั้ง โปรแกรม ENDNOTE

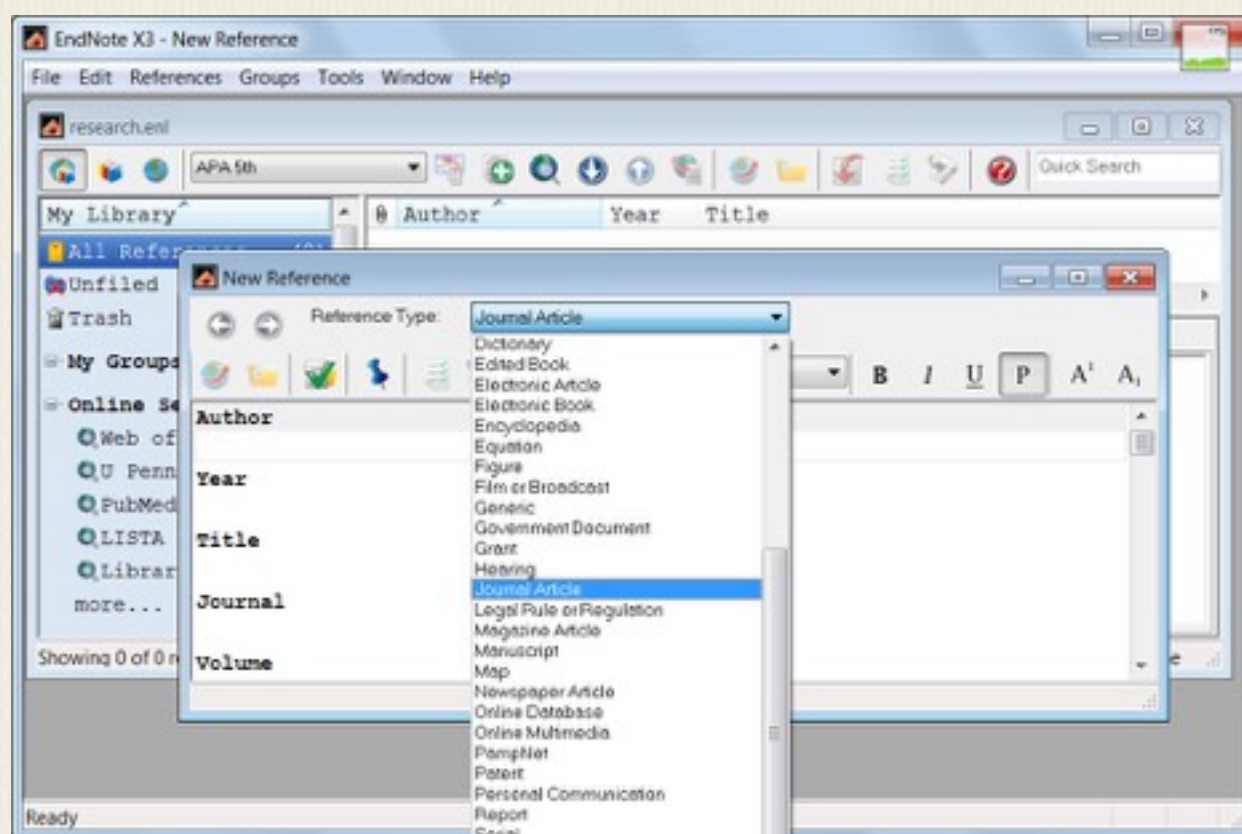
ให้เข้าไปโหลด โปรแกรมได้ที่ <http://www.car.chula.ac.th/topic/ref-mgt> เนื่องจาก
จุฬาฯ ได้ซื้อสิทธิ์การใช้โปรแกรม Endnote X7 มาให้อาจารย์และนิสิตได้ใช้ จึงต้องขอ
password จากเจ้าหน้าที่ตามที่ระบุไว้ในหน้าเว็บ (โปรแกรมนี้ให้ใช้สำหรับอาจารย์และ
นิสิตจุฬาเท่านั้น ห้ามมิให้เผยแพร่แก่บุคคลภายนอก และ password จะเปลี่ยนทุกภาค
การศึกษา)

การตั้งฐานข้อมูลใหม่

การตั้งฐานข้อมูลใหม่

ให้เลือก File – New ตั้งชื่อตามที่ตัวเองต้องการ เช่น research การสร้างข้อมูล
บรรณานุกรมใหม่

เลือก References – New Reference แล้วเลือกประเภทงานที่ตรงกับงานชิ้นนั้น เช่น
เป็น Journal Article, Book, Edited Book, etc. แล้วกรอกข้อมูลใน field ต่างๆ เข้าไปเอง

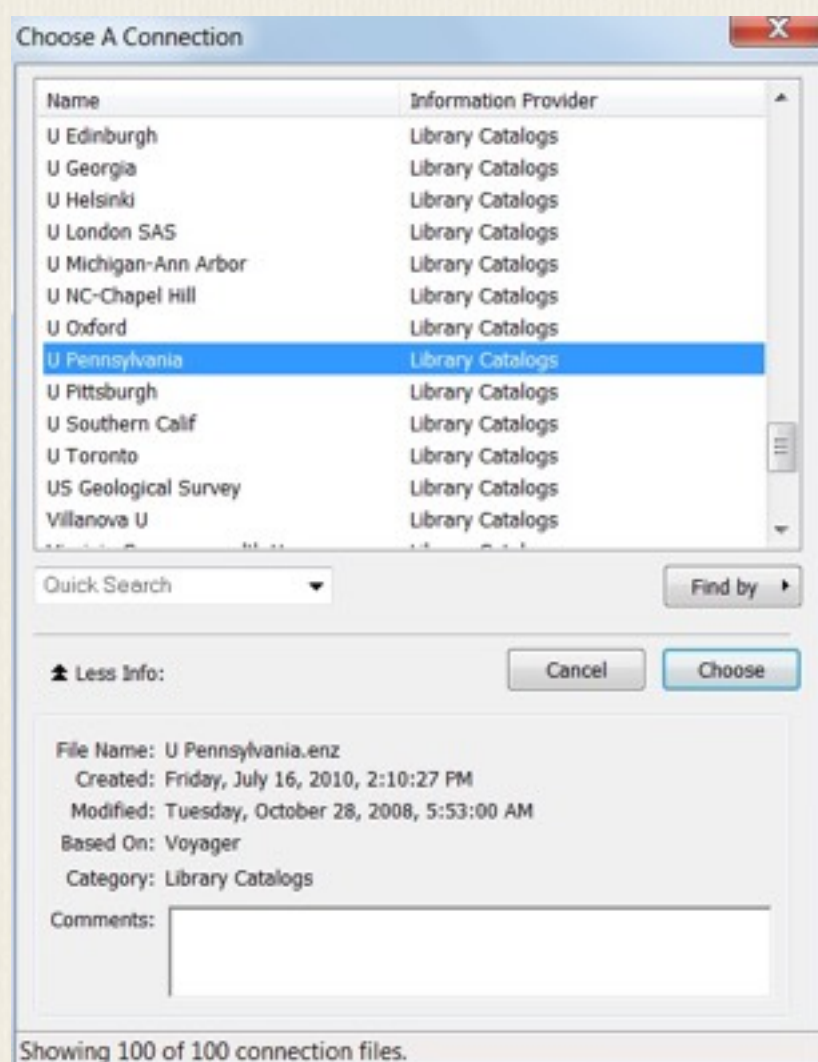


กรอกข้อมูลต่างๆ ใน field ที่เห็น หากเลือกประเภทรายการนั้นผิด ก็สามารถเลือก
เปลี่ยนประเภทใหม่ได้ภายหลัง ข้อมูลใน field ต่างๆจะไปอยู่ใน field เดียวกันของ
ประเภทรายการใหม่

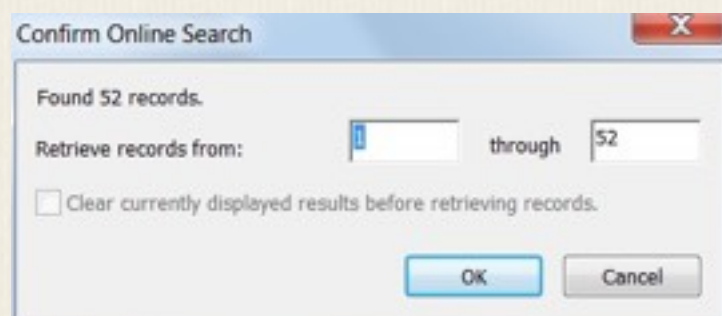
การคัดลอกข้อมูลบรรณานุกรมจากห้องสมุดออนไลน์

การคัดลอกข้อมูลบรรณานุกรมจากห้องสมุดออนไลน์

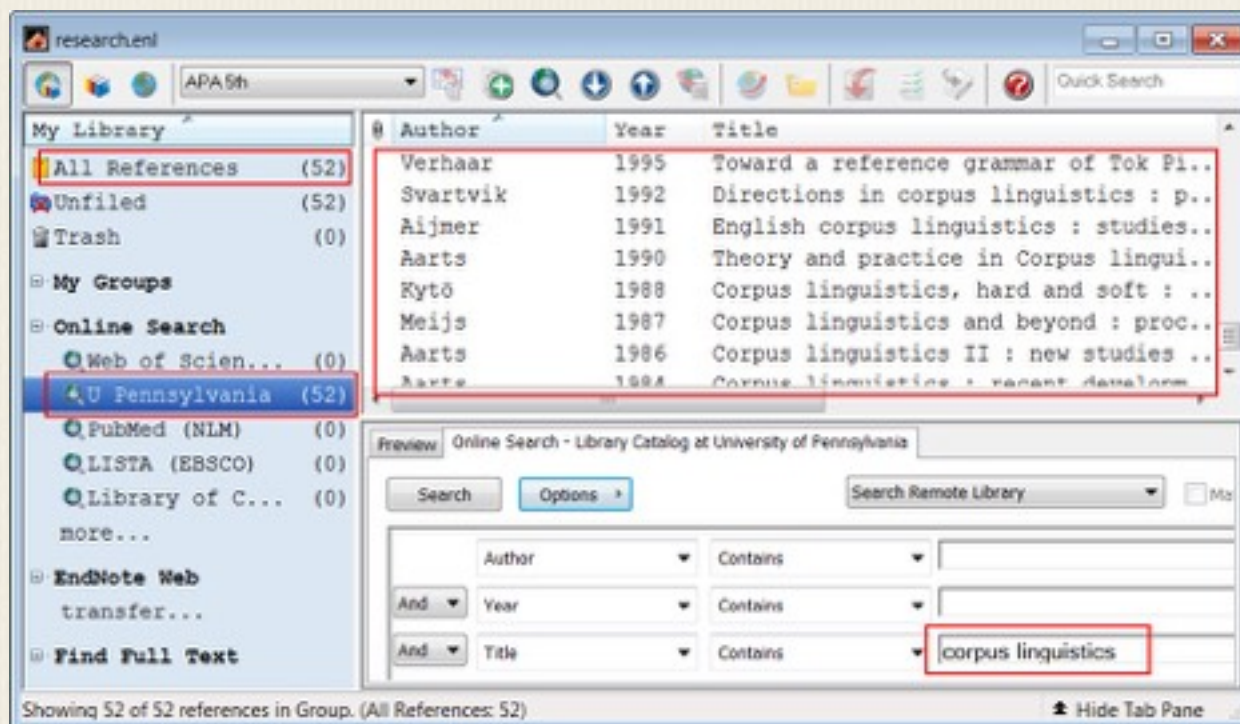
เลือก Tools – Online Search จะเห็นว่ามี online library ที่เราสามารถระบุให้โปรแกรม EndNote เข้าไปค้นรายการในห้องสมุดนั้นได้



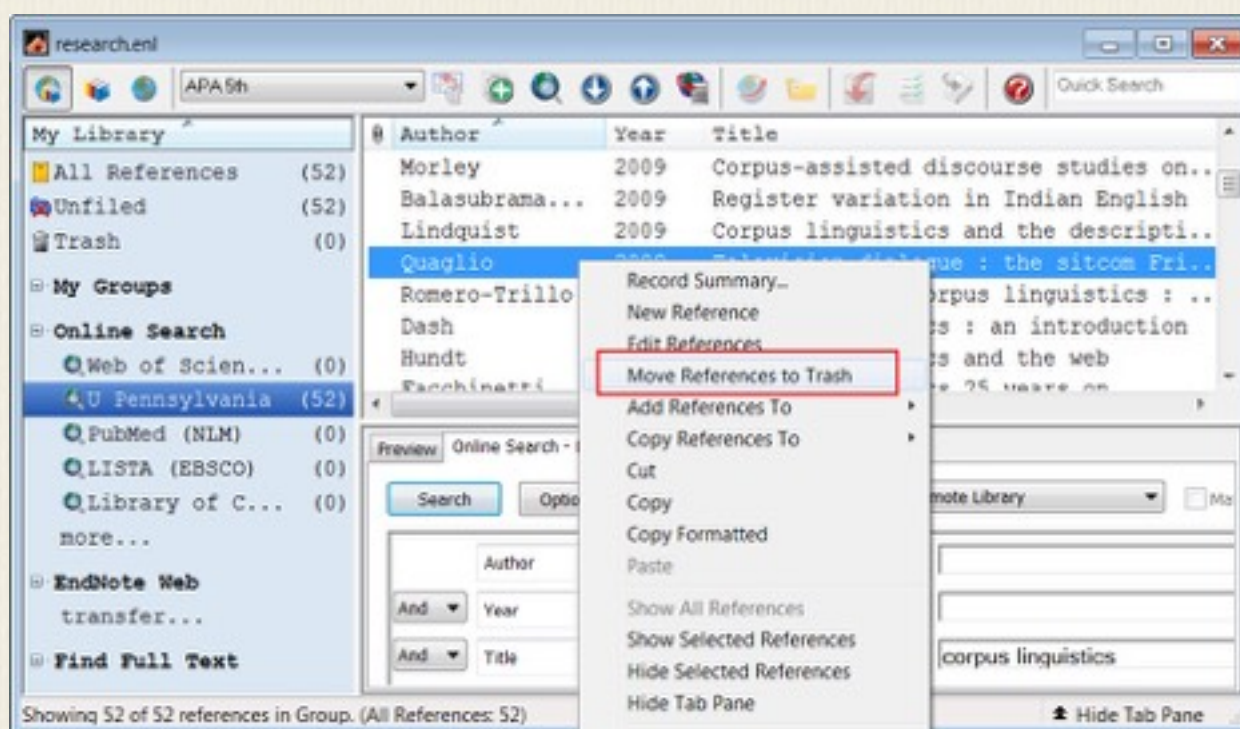
เมื่อเชื่อมต่อได้แล้ว ให้เลือกค้นรายการตามปกติ จะค้นตามชื่อผู้แต่ง ชื่อเรื่อง หรือ คำสำคัญก็ได้ หากค้นพบรายการที่ต้องการจะมีหน้าจอบอกจำนวนรายการที่ค้นได้ออกมา



รายการที่ค้นมาได้จะถูกเพิ่มเข้าไปในฐานข้อมูลปัจจุบันที่ใช้อยู่



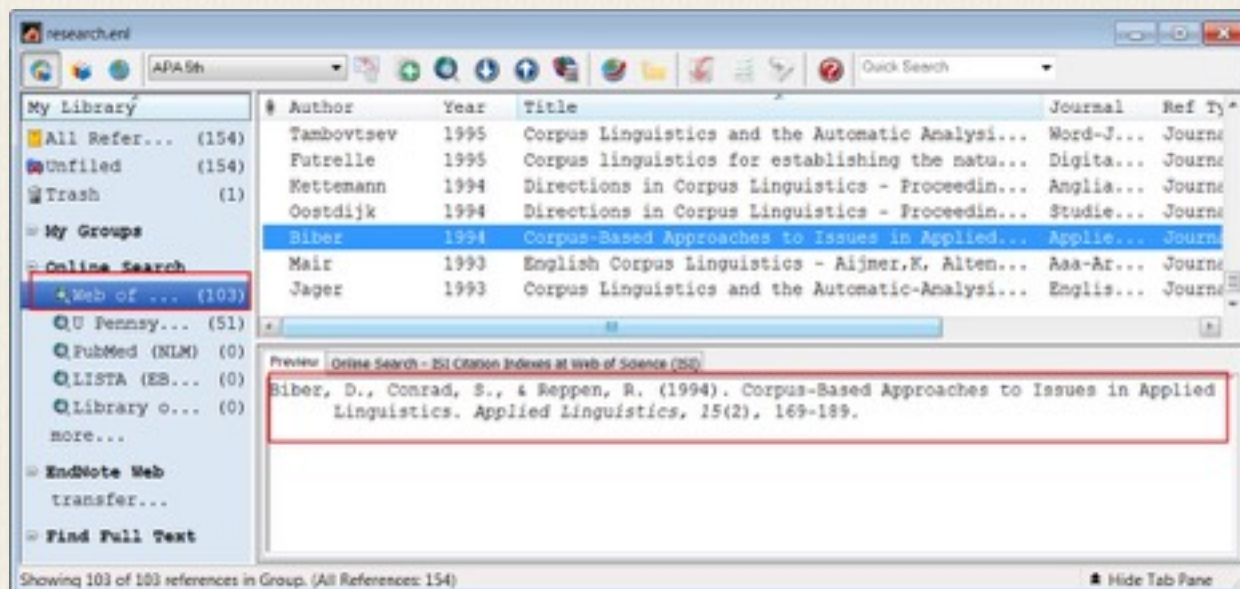
สามารถลบรายการที่ไม่ต้องการจากฐานทิ้งด้วยการเลือกย้ายรายการนั้นลงขยะ หรือกดปุ่ม Ctrl+D หรือ DEL



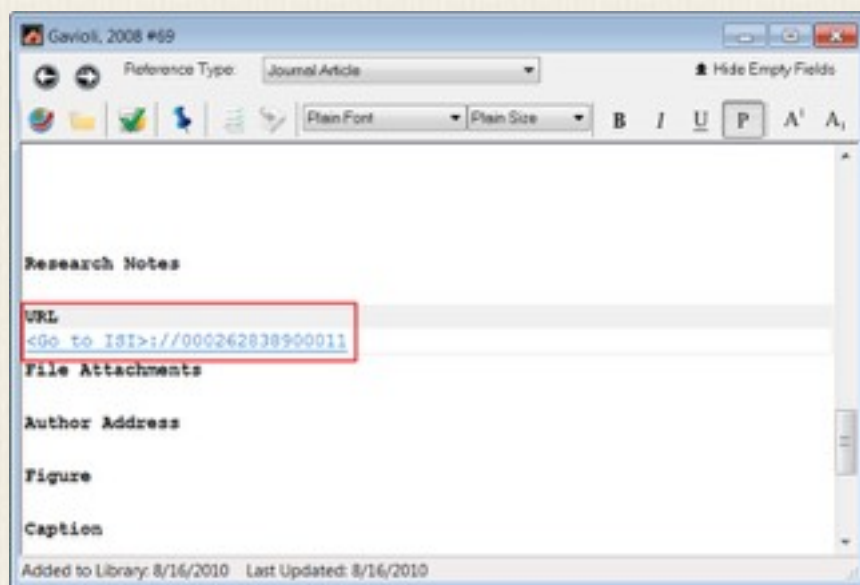
สามารถจัดเรียงข้อมูลตาม field ต่างๆได้ โดยการกดที่ชื่อ field นั้น เช่น กด title จะเป็นการเรียงตามชื่อเรื่อง

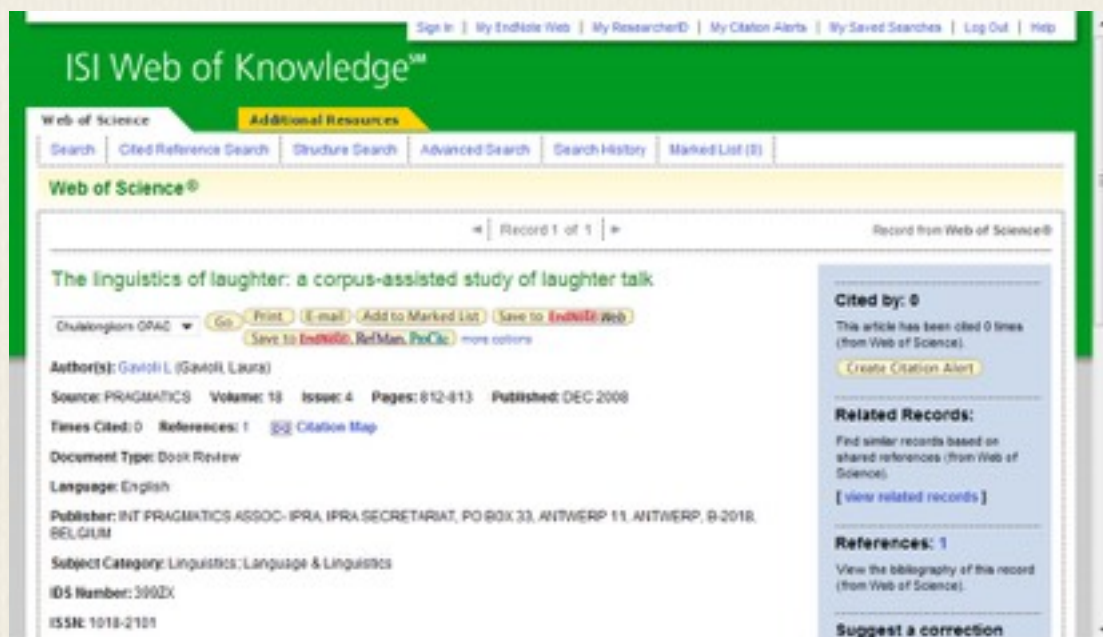
#	Author	Year	Title	Journal	Ref Ty
1	Parodi	2010	Academic and professional discourse genres i...	Studie...	Book
2	Aijmer	2004	Advances in corpus linguistics : papers from...	Langua...	Book
3	Yamamoto	1999	Animacy and reference : a cognitive approach...	Studie...	Book
4	Connor	2004	Applied corpus linguistics : a multidimensio...	Langua...	Book
5	Renouf	2006	The changing face of corpus linguistics	Langua...	Book
6	Murphy	2010	Corpus and sociolinguistics : investigating ...	Studie...	Book
7	Facchinetti	2007	Corpus linguistics 25 years on	Langua...	Book
8	Dash	2008	Corpus linguistics : an introduction	Langua...	Book

หากต้องการค้นหนังสือเราสามารถเลือก Tools - Online Search แล้วเลือกห้องสมุดต่างๆ ที่เปิดให้บุคคลภายนอกค้นผ่านอินเทอร์เน็ตได้ แต่หากต้องการค้นบทความในวารสาร ให้ไปค้นจากฐานข้อมูลที่มีมหาวิทยาลัยบอกรับไว้ เช่น Web of Sciences (ISI) ก็จะได้ รายการบทความต่างๆ มาเก็บใน EndNote ทันที

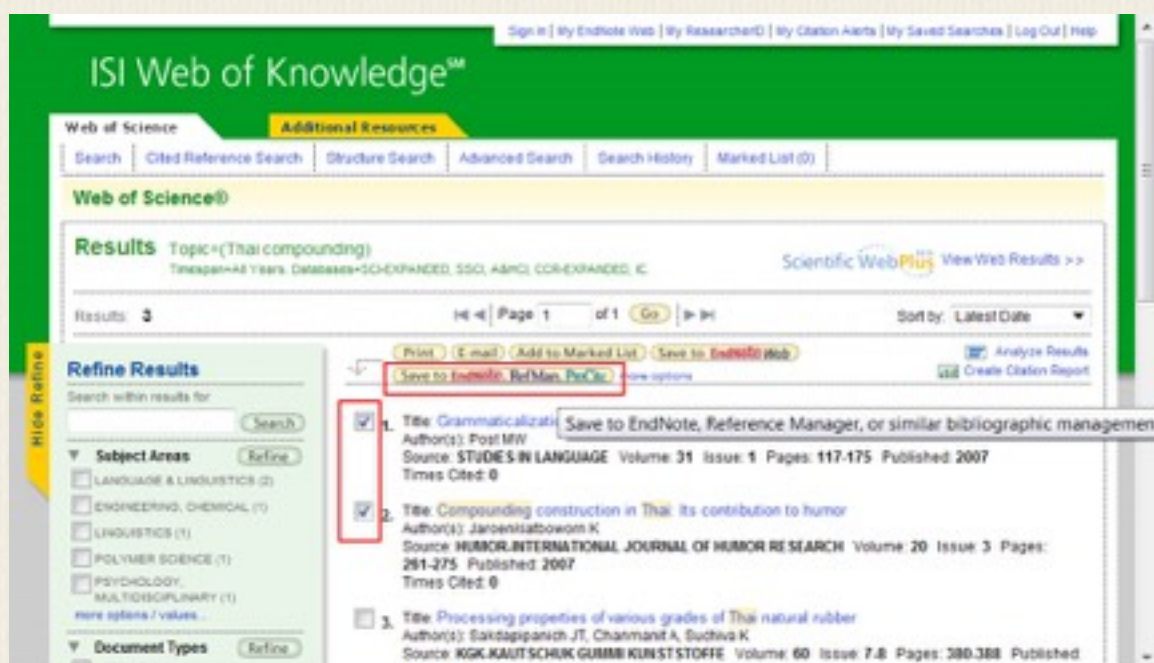


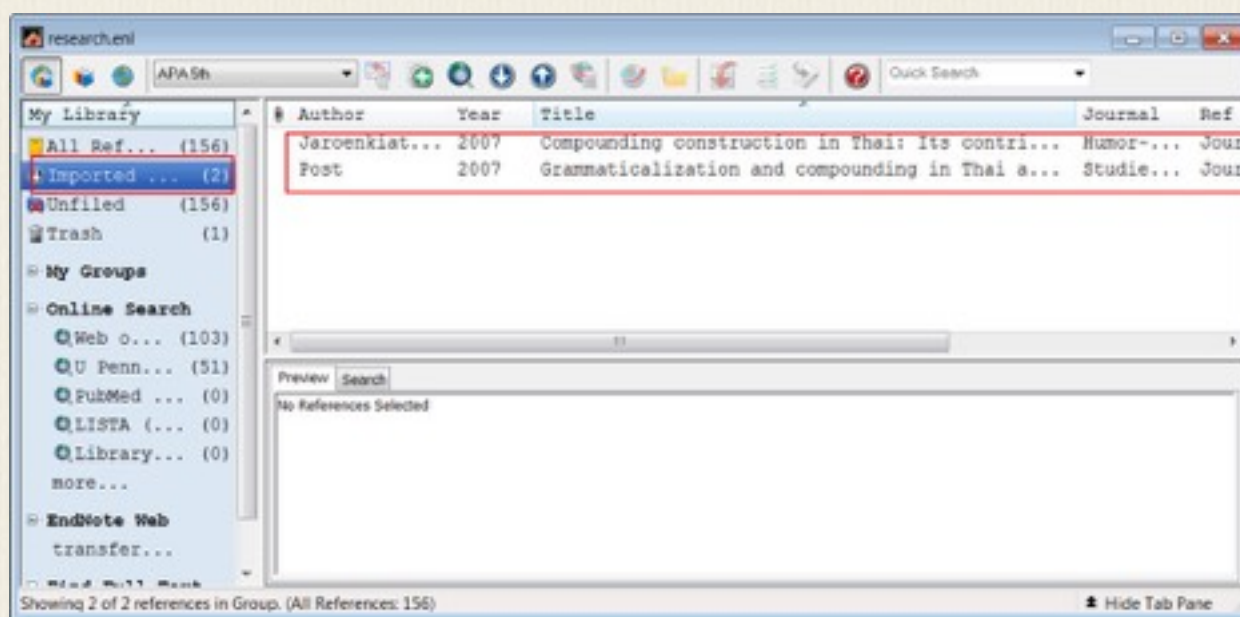
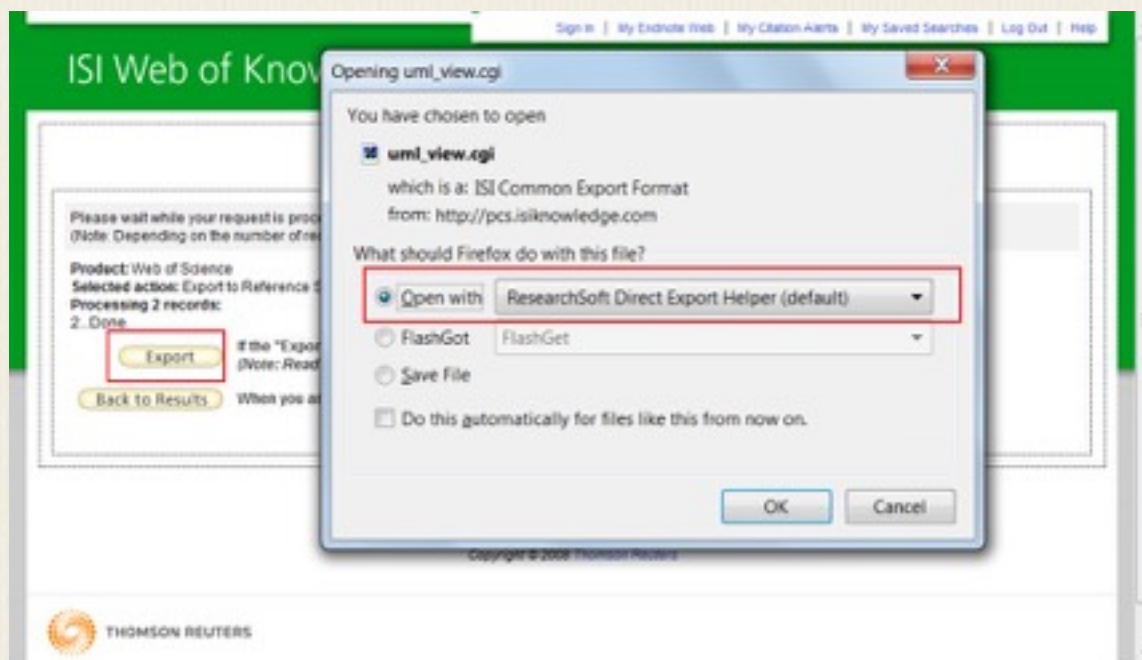
หาก click เข้าไปดูในแต่ละรายการ เราจะเห็นว่ามี link ที่จะโยงเราไปยังเว็บ ISI ได้เลย



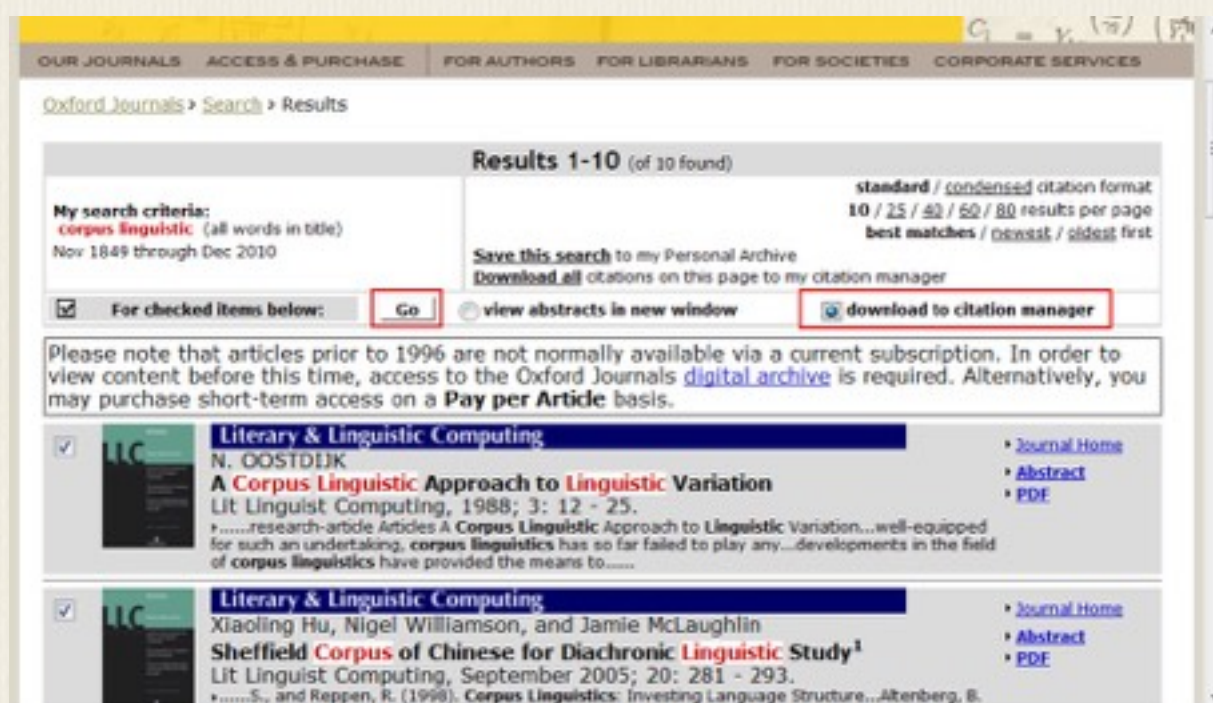


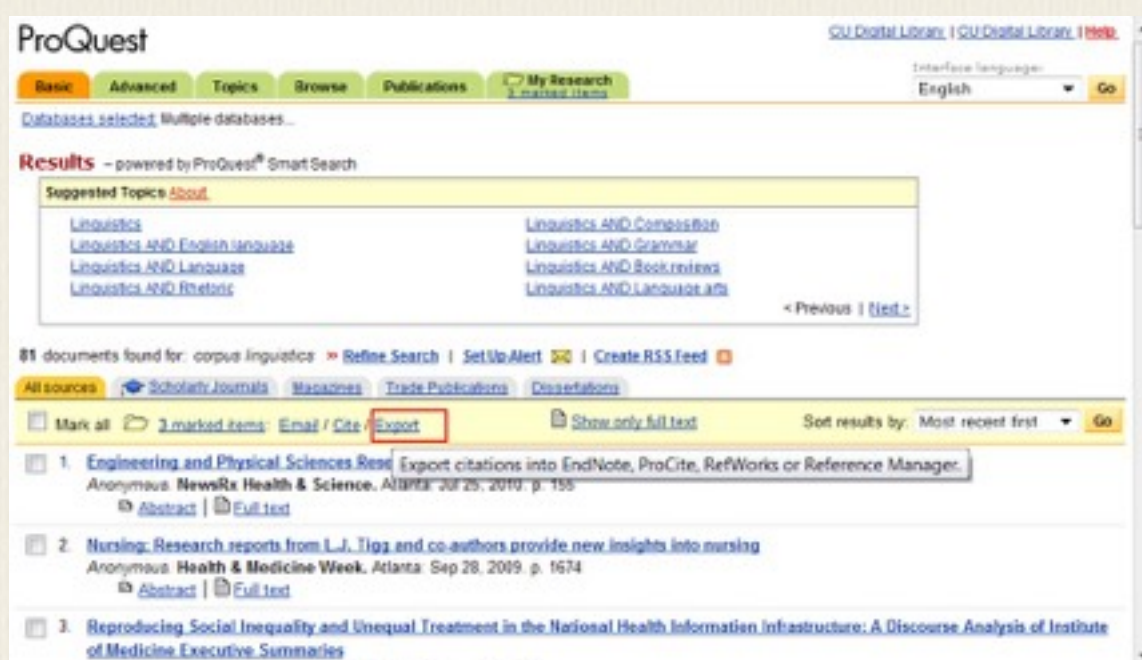
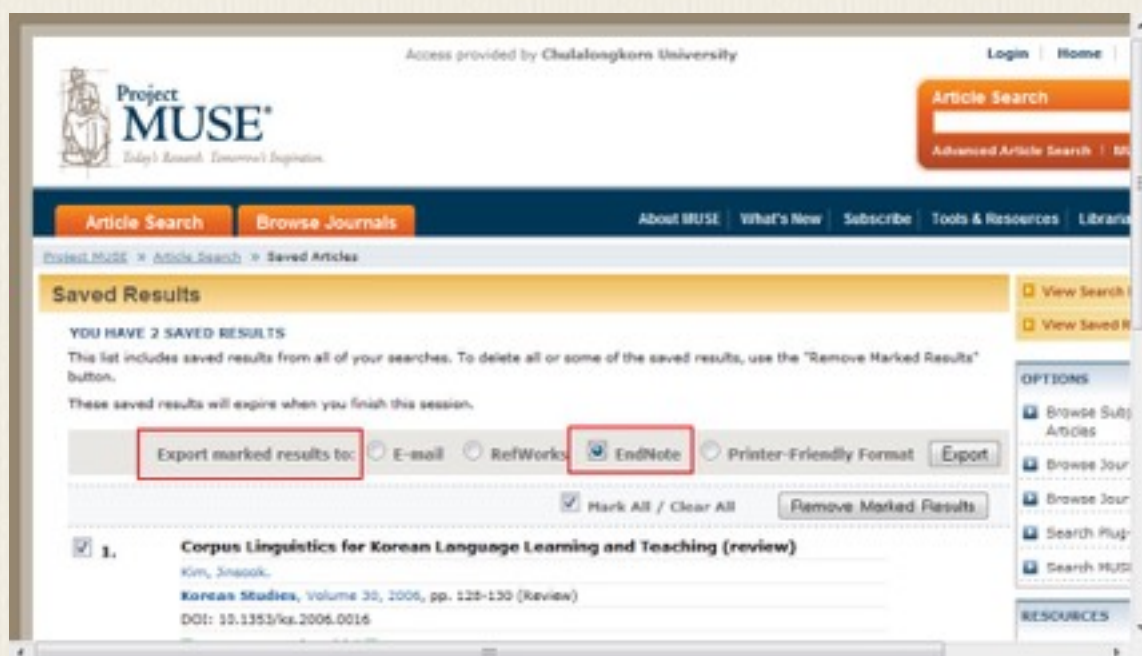
ในฐานะข้อมูลบางฐาน อย่างเช่น ISI จะมีตัวเลือกให้เรา export ข้อมูลมา EndNote ได้ หลังจากค้นข้อมูลที่ต้องการได้แล้ว ให้ mark เลือกรายการที่ต้องการ export แล้วให้ open ด้วย ResearchSoft Direct Export ข้อมูลจะถูกนำเข้าไปใน EndNote ในฐานะข้อมูลที่กำลังเปิดอยู่ทันที





ฐานข้อมูลต่างๆ ตลอดจนวารสารต่างๆ ที่สามารถสืบค้นได้ มักจะมีทางเลือกให้เรา export ข้อมูลบรรณานุกรมออกมาเข้า EndNote ได้ ตัวอย่างเช่น



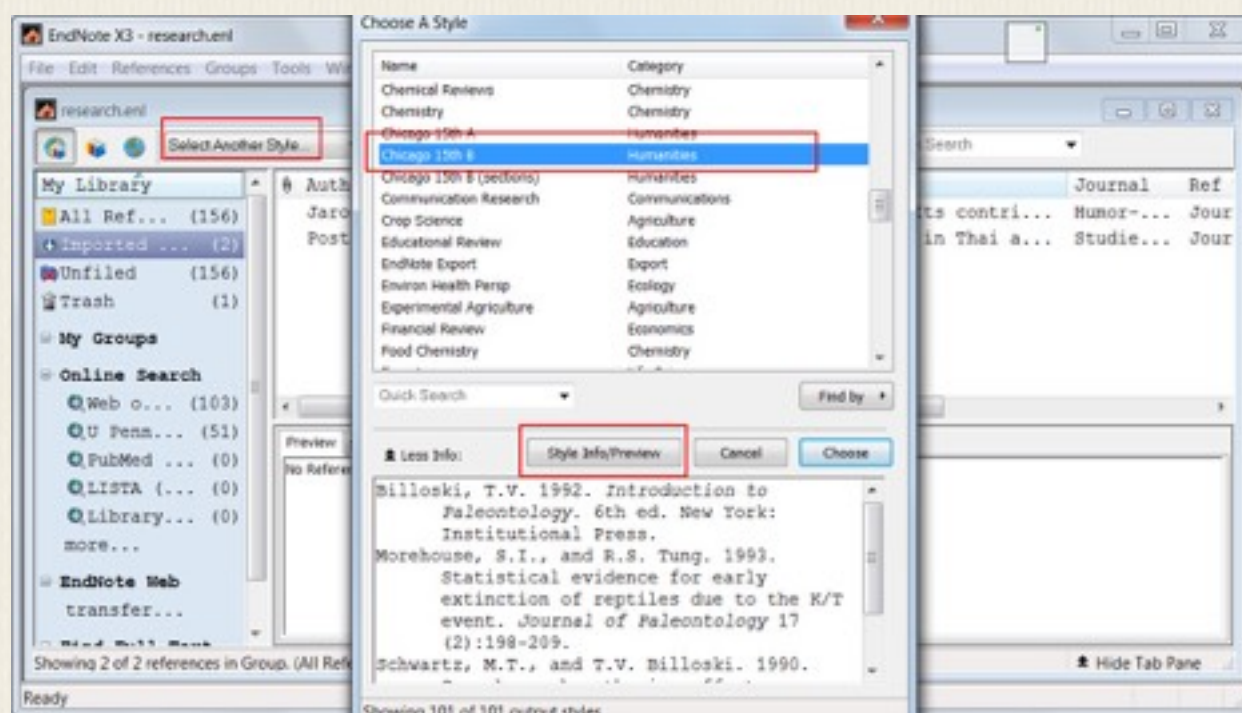


กรณีที่ export เข้า EndNote แบบอัตโนมัติไม่สำเร็จ ให้เลือก save เป็นไฟล์ไว้ไฟล์ที่ได้
ควรมีนามสกุล .enw ถ้าไม่ได้ไฟล์แบบนี้ให้ลองเปลี่ยนนามสกุลเอาเองเป็น .enw เมื่อเปิด
ไฟล์นี้ ข้อมูลจะถูกนำเข้า EndNote ให้เอง

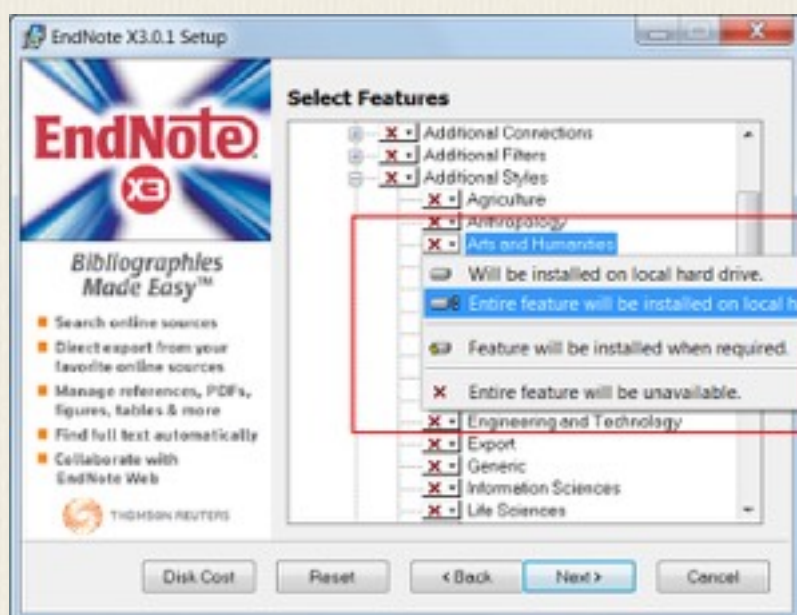
การเลือกแบบบรรณานุกรม

การเลือกแบบบรรณานุกรม

เราสามารถเลือกแบบบรรณานุกรมที่เราต้องการได้ ขึ้นอยู่กับว่าต้องการทำรายการอ้างอิงเพื่อตีพิมพ์ในวารสารใด เช่น Language, IEEE Proceedings ฯลฯ ให้เลือก select another style แล้วเลือกรายการแบบที่ต้องการ กดปุ่ม Style Info/Preview เพื่อดูตัวอย่างที่ออกมาว่าใช่แบบที่ต้องการหรือไม่

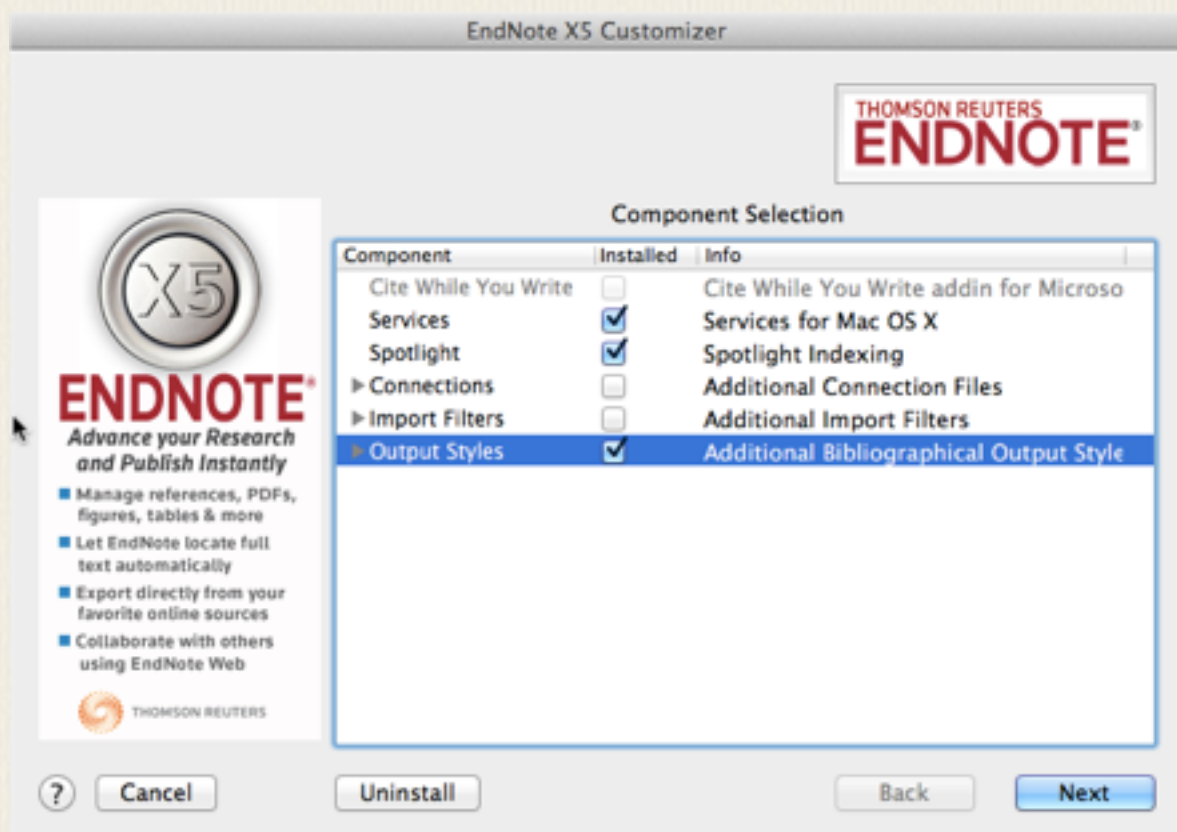


จำนวน style ที่มาเริ่มต้นมี 101 แบบ หากต้องการติดตั้ง style ใหม่เพิ่ม ให้เข้าไปที่ Control Panel ไปในส่วนการ uninstall โปรแกรม เลือก EndNote แล้วเลือก modify โปรแกรม และ add style ในสาขาที่ต้องการเพิ่ม

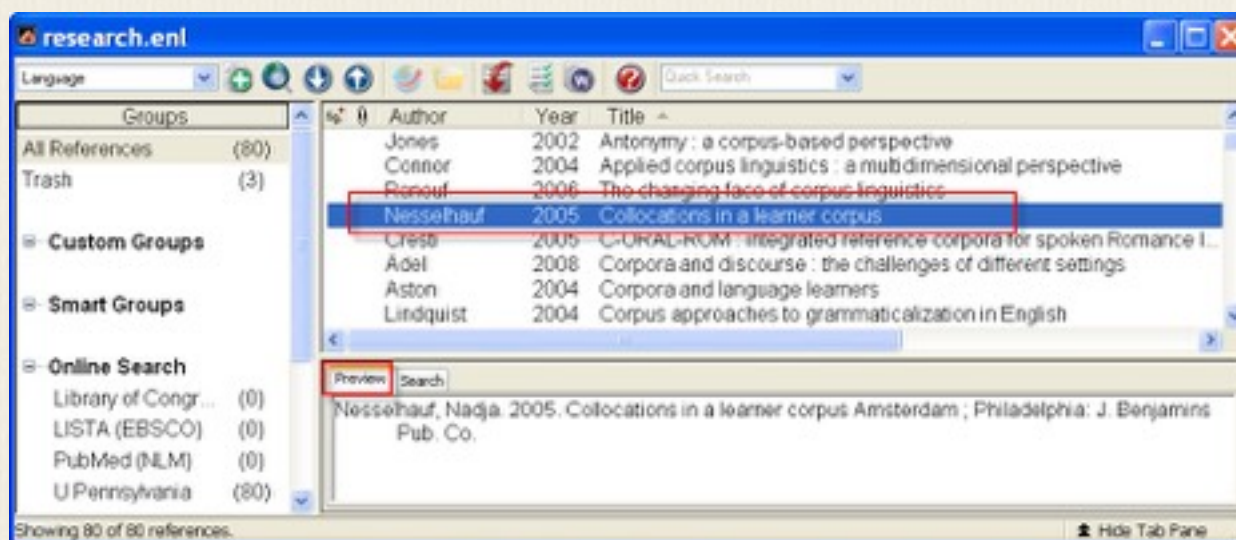


ในระหว่างการติดตั้งนี้ โปรแกรมจะถามหาไฟล์ที่ใช้ในการติดตั้งอย่าง ENX3Inst.msi ซึ่งอาจถูกลบทิ้งไปหลังการติดตั้ง หากเป็นเช่นนั้น ให้ install โปรแกรม EndNote ใหม่ และในระหว่างการติดตั้งนี้ ให้ดูว่า temporary folder อยู่ที่ไหน ให้เข้าไป copy ไฟล์นี้ออกมาก่อนที่จะถูกลบทิ้ง จากนั้นจึงเริ่มขั้นตอนการ modify โปรแกรมใหม่

สำหรับ EndNote X5 (Mac) ให้เลือก EndNote - Customizer จะมีหน้าจอขึ้นให้เลือกเพิ่ม Output Style



สไตส์ต่างๆ ที่เลือกใช้สามารถ preview ดูได้ว่าใช่แบบการอ้างอิงที่ต้องการหรือไม่ โดยเลือกแท็บ preview จะเห็นรายการสิ่งที่ถูกเลือกไว้



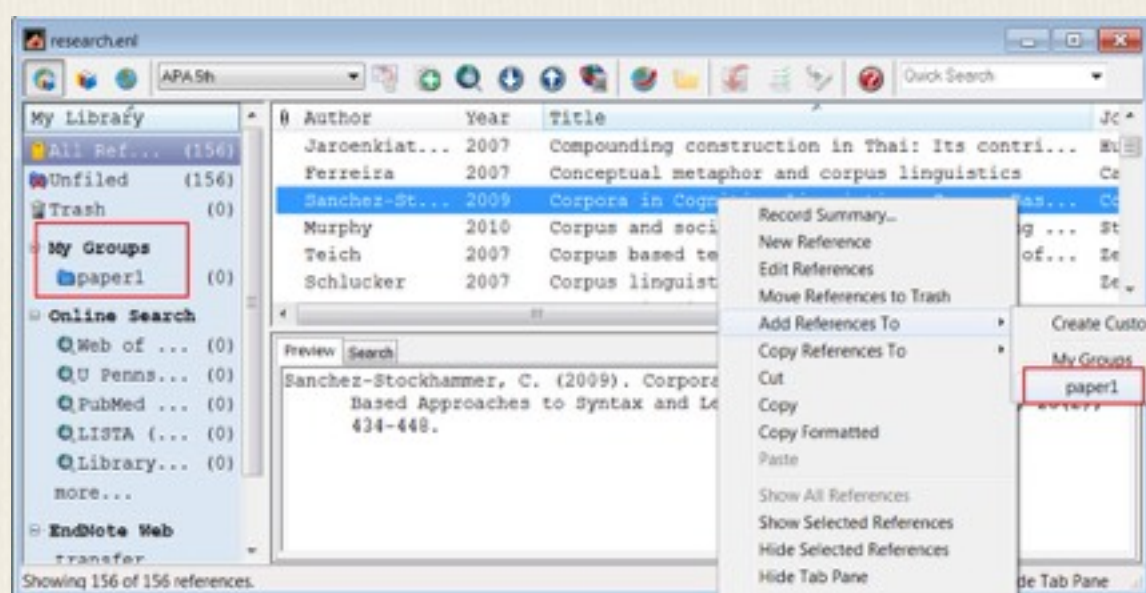
การสร้างกลุ่มรายการอ้างอิง

การสร้างกลุ่มรายการอ้างอิง

ในการเขียนแต่ละบทความเรามักจะอ้างอิงเฉพาะบางรายการที่เกี่ยวข้อง เราจึงควรสร้าง group สำหรับแต่ละงานขึ้นมา รายการบรรณานุกรมที่อยู่ใน reference สามารถ copy ไปยัง group ต่างๆ ได้

เลือก Groups – Create Group จะเห็นว่าเราสามารถใส่ชื่อ group ใหม่ได้ เช่น ใส่ว่า paper1 สำหรับเก็บ references ที่ใช้ในการเขียน paper1

เลือกใส่รายการที่ต้องการจาก main references โดยเลือกรายการที่ต้องการ กดเมาส์ปุ่มขวา เลือก Add reference to – paper1 จะเห็นว่า paper1 จะมีจำนวนรายการเพิ่มขึ้นอีก 1 ทำแบบนี้กับทุกรายการที่ต้องการใช้สำหรับ paper1 (สามารถ add หลายรายการพร้อมกันก็ได้ ให้เลือกทีเดียวหลายรายการ กด Ctrl+ mouse click)



รายการที่เลือก add เข้า group ก็ยังคงอยู่ใน All references เราสามารถลบรายการที่ไม่ต้องการใน group ออกได้ โดยไม่มีผลไปการลบรายการออกจากรฐานข้อมูลหลัก ถ้าต้องการลบจากรฐานข้อมูลหลักต้องเข้าไปที่ All references แล้วเลือกรายการที่ต้องการลบ

Group ใดที่เมื่อเลิกใช้แล้ว ก็สามารถลบทิ้งได้ เพราะจะไม่มีผลต่อรายการที่เก็บใน
ฐานข้อมูลหลัก (All References)

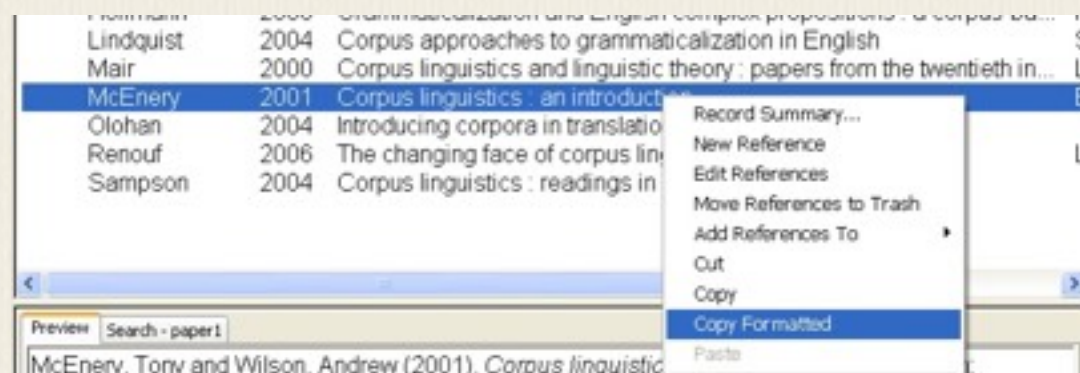
การสร้างรายการบรรณานุกรม

การสร้างรายการบรรณานุกรม

เลือก group ที่ต้องการสร้างรายการบรรณานุกรม เช่น paper1

เลือก Edit – Select All และ Edit – Copy Formatted จากนั้นไปยัง paper ที่เราเขียนอยู่ paste ลงตำแหน่งที่ต้องการ

ถ้าต้องการเลือกเฉพาะบางรายการ กด right-click ที่รายการนั้น เลือก Copy Formatted



การเลือกแต่ละ style จะให้ format ที่แตกต่างกัน เช่น ถ้าเลือกวารสาร Language in Society จะได้

McEnery, Tony and Wilson, Andrew (2001). *Corpus linguistics : an introduction*. Edinburgh: Edinburgh University Press.

ถ้าเลือกวารสาร language จะได้

McEnery, Tony & Andrew Wilson. 2001. *Corpus linguistics : an introduction* Edinburgh: Edinburgh University Press.

เมื่อเขียนบทความเสร็จและได้รายการบรรณานุกรมที่เกี่ยวข้องลงใน group แล้ว ให้ไปที่ group นั้น เลือก Edit-Select All แล้วเลือก Edit-Copy Formatted จากนั้นไปที่ไฟล์ MS Word ของบทความนั้น แล้ว paste ลงในส่วนบรรณานุกรมของบทความก็จะได้รายการทั้งหมดใน format ที่ถูกต้องทันที

การใช้งานร่วมกับ โปรแกรม Word

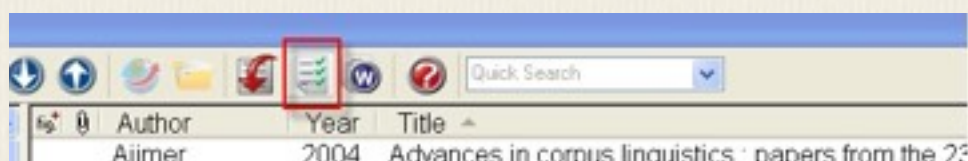
การใช้งานร่วมกับ โปรแกรม WORD

เราสามารถใส่รายการอ้างอิงและบรรณานุกรมไประหว่างการทำงานได้ที่ละรายการ โปรแกรม EndNote version นี้สามารถทำงานร่วมกับ MS Word 2003 และ 2007 ได้

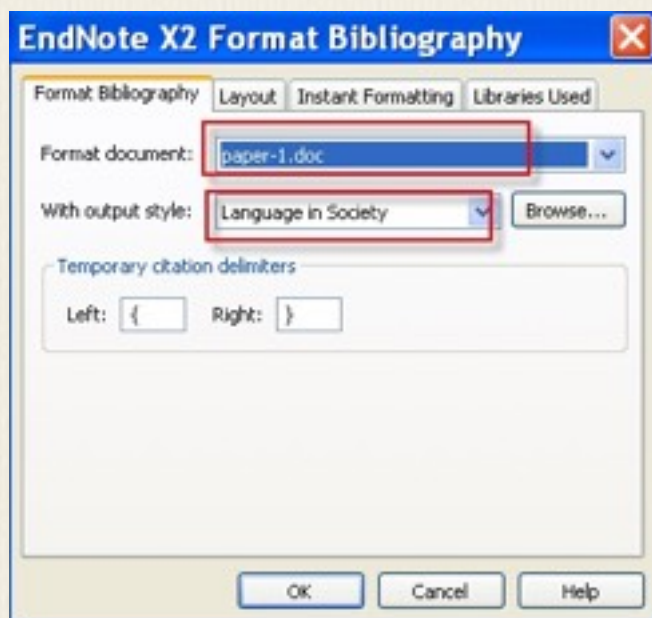
ผู้ที่ใช้ Office 2010 จะพบว่าเมื่อติดตั้ง โปรแกรม EndNote แล้ว MS Word จะ hang หลังจากเปิดทำงาน ทั้งนี้เพราะ macro จาก โปรแกรม EndNote วิธีแก้ไขให้เปิด Word โดยไม่เรียก macro โดยเรียกจาก command line “winword /a” เมื่อเข้าโปรแกรม Word 2010 แล้ว ไปที่ File – Options – Add-Ins แล้วเลือก Manage Com Add-Ins ในรายการ macro ที่มาจาก EndNote ให้ remove macro นั้นออกไป จากนั้นจึงจะเปิด Word 2010 ได้ตามปกติ

คำอธิบายต่อไปนี มาจากตัวอย่างการใช้กับ Word 2007

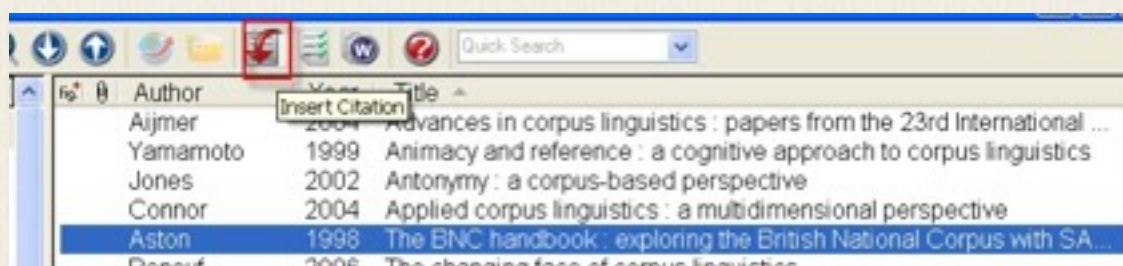
เลือก Format Bibliography



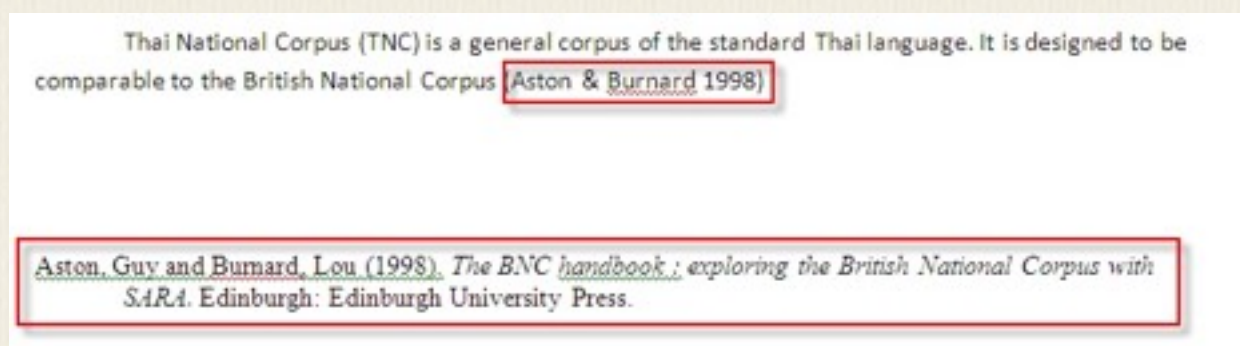
จะได้หน้าจอที่ให้เราเลือกไฟล์ Word ที่เราต้องการทำงาน และเลือกแบบบรรณานุกรมที่ต้องการ



ในระหว่างการพิมพ์งานเราสามารถ insert reference ไปได้เลย ทั้งตัว reference และ รายการ reference จะถูกนำไปใส่ในไฟล์ Word โดยอัตโนมัติ โดยเลือกรายการที่ต้องการ แล้วเลือกปุ่ม Insert Citation



Citation และ reference ตาม format style ที่เราเลือกจะถูก paste ลงในไฟล์ Word ที่เรา ระบุไว้ใน Format Bibliography



เมื่อพิมพ์งานต่อไปแล้วมีการอ้างอิงรายการเพิ่มก็ทำได้เรื่อยๆ จนจบบทความ รายการอ้างอิงต่างๆ ก็จะถูกนำมาใส่ในบทความนั้น ตรงตาม format ที่เราได้เลือกไว้

Thai National Corpus (TNC) is a general corpus of the standard Thai language. It is designed to be comparable to the British National Corpus (Aston & Burnard 1998) in terms of its domain and medium proportions. However, only written texts are collected in the TNC, and the corpus size is targeted at eighty million words. In addition to domain and medium criteria, texts are also selected and categorized on the basis of their genres. We adopted Lee's idea of categorizing texts into different genres based on external factors like the purpose of communication, participants, and the settings of communication (Lee 2002)

|

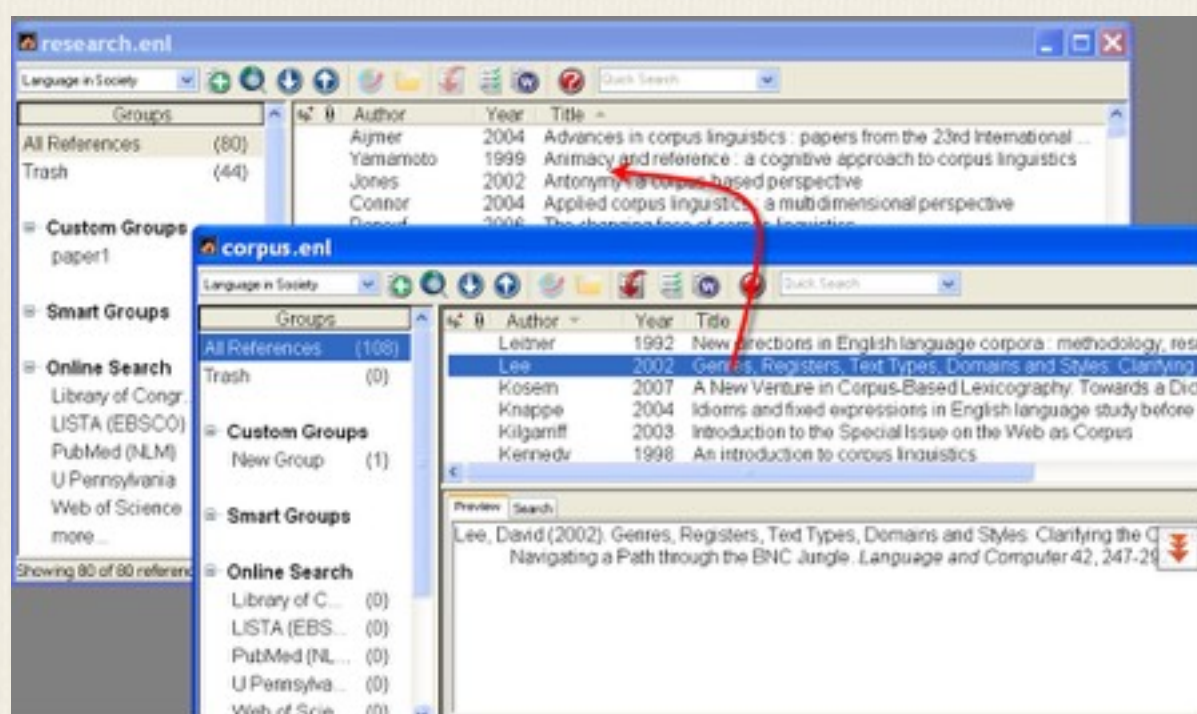
Aston, Guy and Burnard, Lou (1998). *The BNC handbook : exploring the British National Corpus with S.I.R.I.* Edinburgh: Edinburgh University Press.

Lee, David (2002). Genres, Registers, Text Types, Domains and Styles: Clarifying the Concepts and Navigating a Path through the BNC Jungle. *Language and Computer* 42, 247-292.

การย้ายข้อมูลระหว่างฐานข้อมูล

การย้ายข้อมูลระหว่างฐานข้อมูล

เราสามารถเปิดฐานข้อมูลมากกว่าหนึ่งฐานขึ้นมาใช้ได้ และสามารถ copy รายการจากฐานหนึ่งไปยังอีกฐานหนึ่งได้ ให้เลือกรายการที่ต้องการ แล้วกดลากไปยังอีกฐานหนึ่ง





โปรแกรม Zotero

โปรแกรม Zotero เป็นโปรแกรมช่วยจัดเก็บข้อมูลบนเว็บที่พบ เดิมใช้กับ Firefox ทำให้สามารถคัดลอกข้อมูลบรรณานุกรมจากเว็บได้โดยตรง และสามารถย้ายข้อมูลไปเข้าใน EndNote ได้ ปัจจุบัน มีทั้งแบบที่เป็น plugin ของ Firefox และแบบที่เป็นโปรแกรมเอกเทศที่เชื่อมโยงกับ browser อื่น คือ Chrome, Safari หากเลือกแบบ โปรแกรม add-on ผู้ใช้จะต้องมีโปรแกรม Firefox ก่อน ซึ่งเป็น web browser ตัวหนึ่ง เมื่อติดตั้ง Firefox แล้วให้ไปที่ <http://www.zotero.org/> เพื่อโหลดโปรแกรม addon Zotero นี้

Zotero สามารถทำสิ่งต่างๆ เหล่านี้ได้ขณะที่ browse internet : ดึงเอา citation ออกมาโดยอัตโนมัติ เก็บไฟล์ pdf, web และรูปภาพ สร้าง library ของตัวเองโดยการเก็บข้อมูลและงานต่างๆ ที่พบระหว่างการใช้อินเทอร์เน็ต

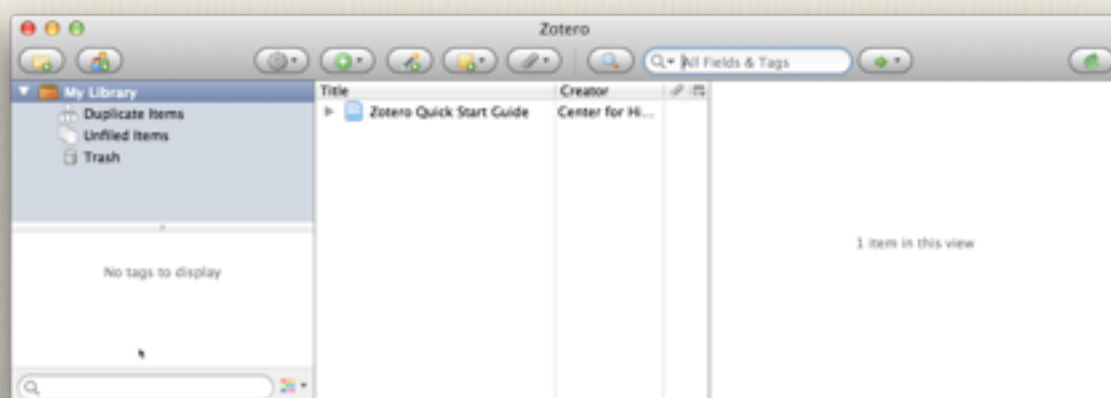
การติดตั้ง โปรแกรม Zotero Standalone

การติดตั้ง โปรแกรม ZOTERO STANDALONE

ในกรณีที่ไม่ต้องการใช้ Firefox แต่จะใช้ Chrome หรือ Safari ไปที่

<http://www.zotero.org/> เลือก version standalone แล้วติดตั้ง โปรแกรม exe (Windows)

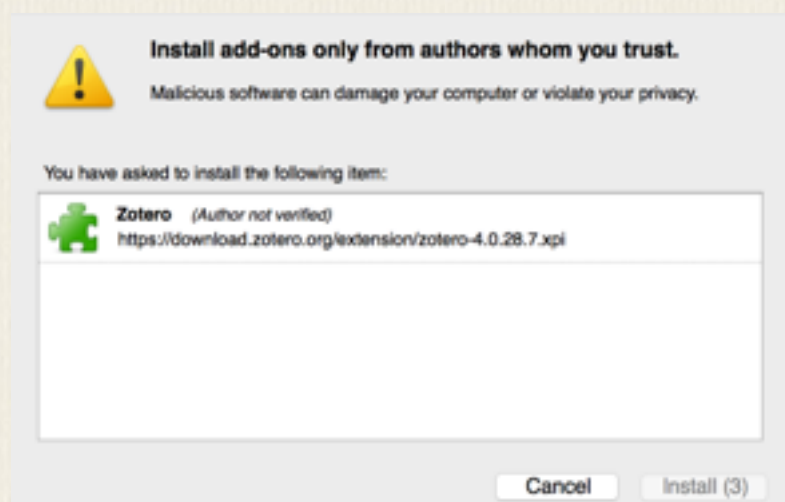
หรือ dmg (Mac OS X) เมื่อเรียก โปรแกรม Zotero ขึ้นมาจะเห็นหน้า โปรแกรมแบบนี้ จากนั้นให้ติดตั้ง Zotero Connectors ซึ่งก็คือ add on โปรแกรมสำหรับ browser ที่ต้องการใช้ แต่ในที่นี้ จะนำเสนอการใช้ Zotero on Firefox เพราะเป็น โปรแกรมหลักที่พัฒนาและมีความสะดวกในการใช้งานดึงข้อมูลต่างๆบน browser มาเก็บไว้



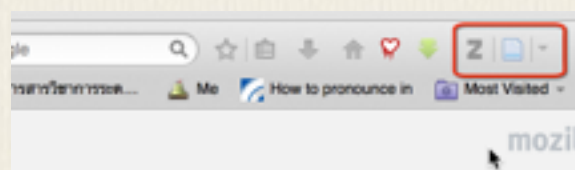
การติดตั้ง โปรแกรม addon Zotero

การติดตั้ง โปรแกรม ADDON ZOTERO

ไปที่ <http://www.zotero.org/> เลือก version ใหม่สุด เมื่อกดแล้วจะได้หน้าจอถามให้ติดตั้งหรือไม่



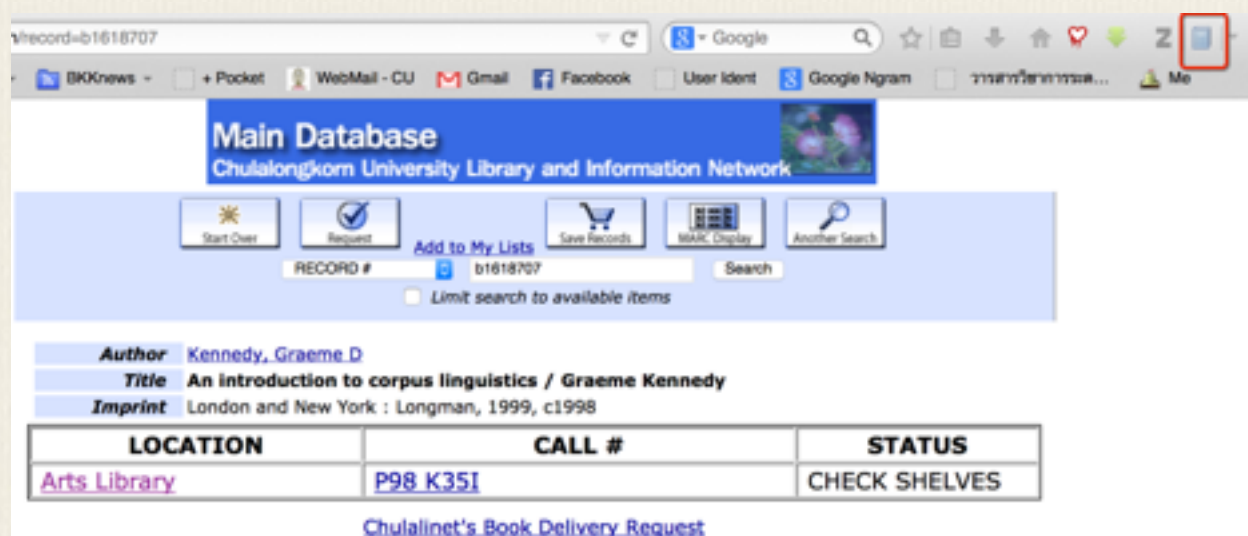
เมื่อติดตั้งแล้วให้ restart โปรแกรม Firefox จะเห็นที่มุมขวามือไอคอนของ Zotero อยู่



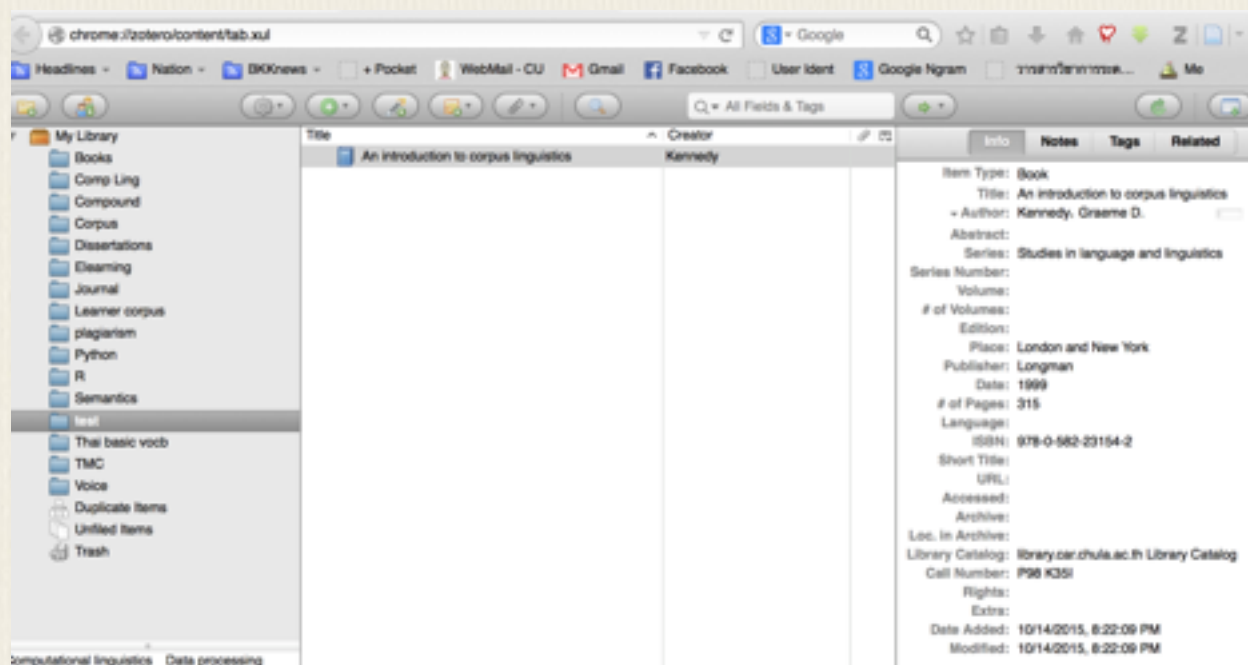
การเก็บรายการข้อมูล

การเก็บรายการข้อมูล

เมื่อเข้าไปยัง web ที่มีรายการหนังสือ เช่น เว็บห้องสมุด เมื่อค้นรายการหนังสือที่ต้องการได้ จะสังเกตว่าที่ address bar ด้านขวาจะมีไอคอนรูปหนังสือ หมายความว่าสามารถนำข้อมูลบรรณานุกรมมาเก็บใน Zotero ได้ ให้กดที่ไอคอนนี้

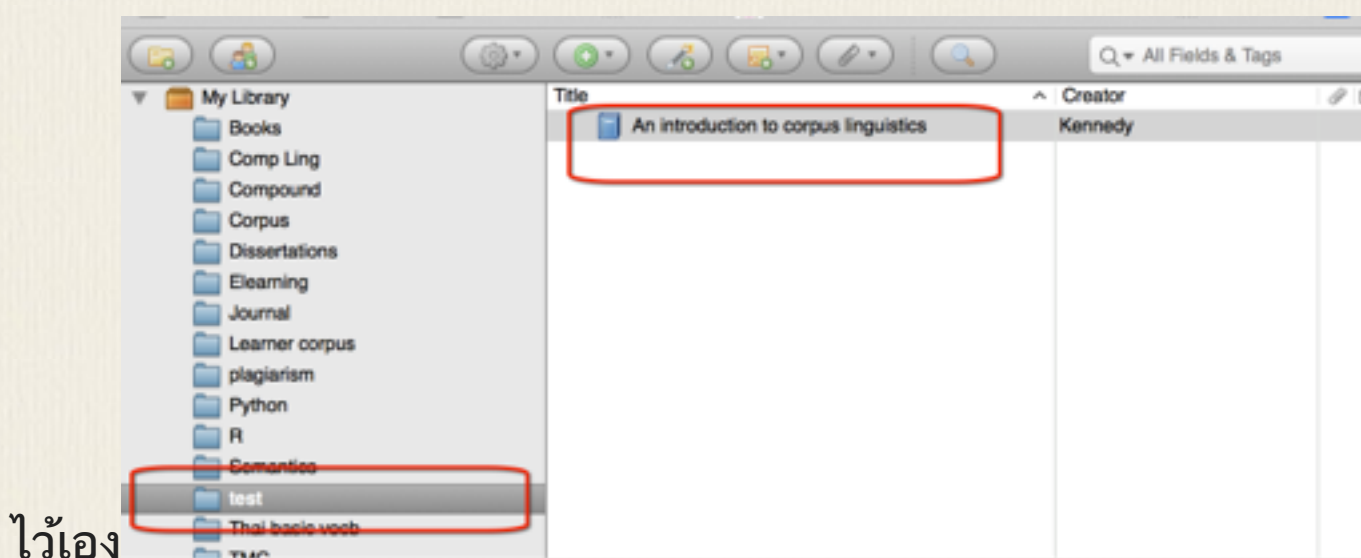


เมื่อกดที่ไอคอน Zotero ตรงมุมขวาของ Firefox จะเห็นหน้าจอ Zotero ขึ้นมาในอีกแท็บ และรายการหนังสือนี้จะถูกเก็บใน Zotero แล้ว



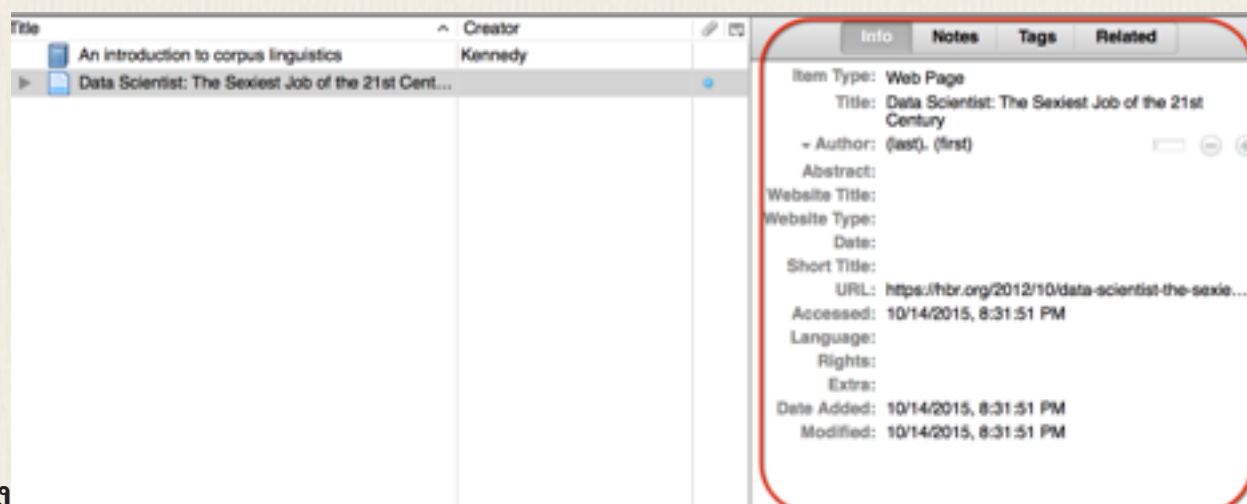
เราควรสร้าง collection เฉพาะสำหรับงานวิจัยแต่ละเรื่อง ที่มุมซ้ายบนของ Zotero มีไอคอน New Collection ให้ตั้งชื่อ collection ของงานตนเอง แล้วย้ายรายการข้อมูลที่ต้องการไปยัง collection ที่สร้างขึ้น

My Library จะเก็บรายการข้อมูลทั้งหมด item เดียวกันสามารถไปอยู่ในหลายๆ collection ได้ collection จึงไม่ได้เก็บข้อมูลรายการจริง แต่คล้ายๆ กับ playlist ใน media player ที่เราเลือกเฉพาะชื่อรายการมาจัดเป็น collection หนึ่งๆ แต่ตัวรายการจริงเก็บไว้อีกที่ เวลาที่ต้องการให้ข้อมูลไปอยู่ใน collection ไหน ก็ให้เปิด Zotero ค้างไว้ที่หน้า collection นั้น เมื่อมีการเก็บรายการข้อมูลใหม่ ชื่อรายการนั้นจะไปอยู่ใน collection ที่เราเลือกค้าง



เว็บที่เกี่ยวข้องกับรายการหนังสือ วารสาร เช่น Amazon เว็บห้องสมุดต่างๆ เมื่อเราค้นเจอรายการที่ต้องการ โดยปกติเราจะเห็นไอคอนรูปหนังสือให้ save รายการเก็บไว้ได้ จึงค่อนข้างสะดวกต่อการเก็บรายการอ้างอิงต่างๆ ไว้ใช้ภายหลัง

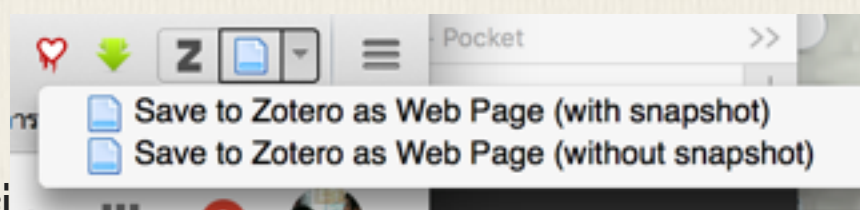
บางเว็บไซต์ เราจะเห็นไอคอนให้ save to Zotero เป็นรูปอื่นไม่ใช่รูปหนังสือ หมายความว่า เราสามารถเก็บข้อมูลเว็บนั้นลงใน Zotero ได้ ซึ่งจะได้ web address และข้อมูลอื่นที่



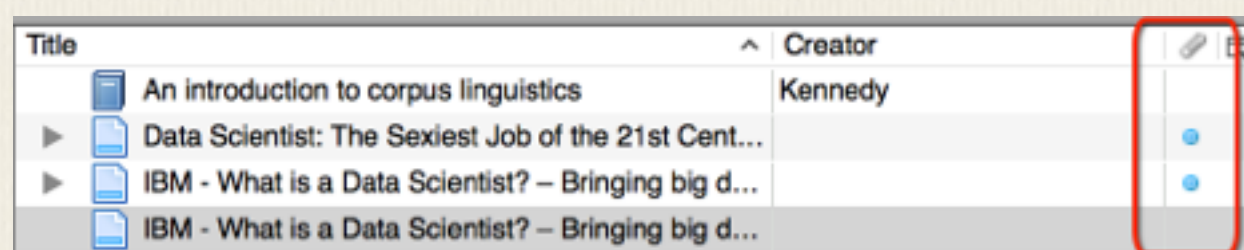
เกี่ยวข้อง

การลบรายการจาก My Library ให้เลือกรายการนั้นแล้วกดปุ่ม DEL ถ้าต้องการเพียงลบออกจาก collection (แต่ยังอยู่ใน My Library) ให้เลือกรายการใน collection นั้น แล้วกด CTRL-DEL

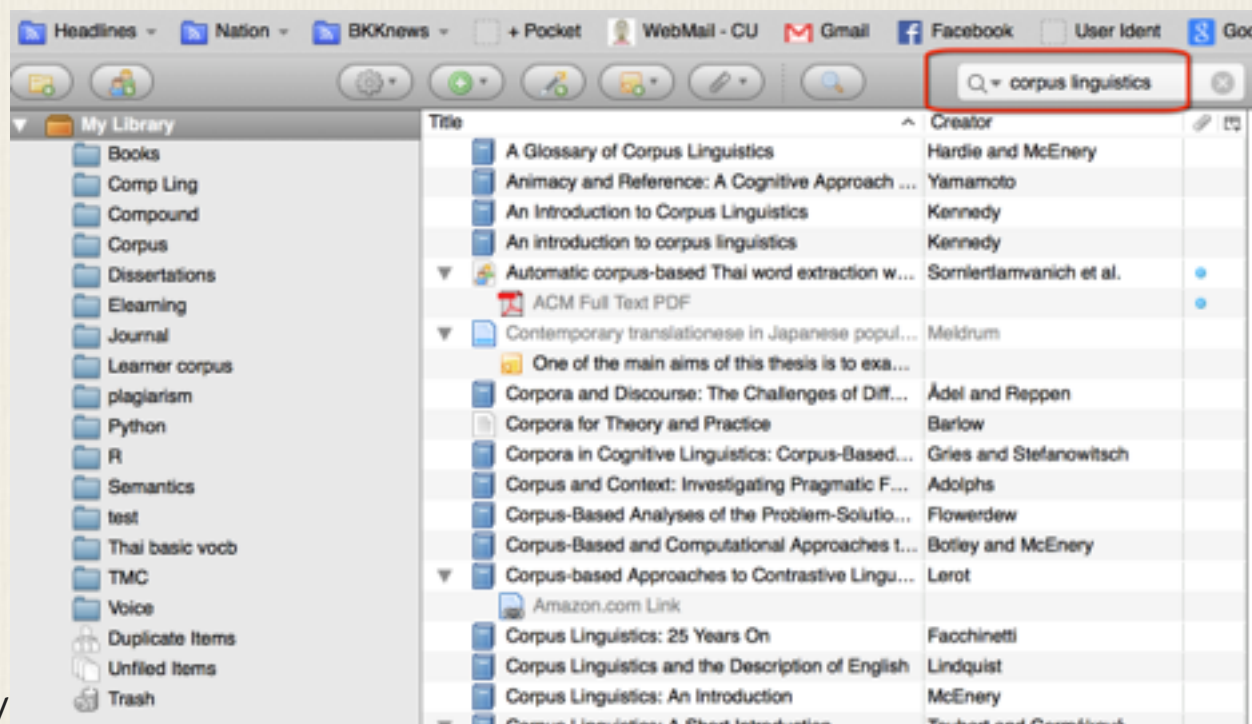
การ save item ลงใน Zotero สามารถเลือกได้ว่าจะ save เนื้อหาหน้าเว็บด้วยไหม (with snapshot) หากดูรายการที่ save ใน Zotero จะเห็นความต่างว่า แบบที่เก็บ snapshot จะมี attachment ติดอยู่ด้วย ข้อมูลที่เก็บแบบ snapshot ว่าจะสามารถมาเปิดดูภายหลังได้



แม้ว่าไม่ได้ต่ออินเทอร์เน็ตอยู่



ในโหมดที่ดู Sanpshot เราจะเห็นว่าข้อมูลนั้นเก็บที่เครื่องเรา ดูได้จาก address จะเป็น



zotero://attachment/

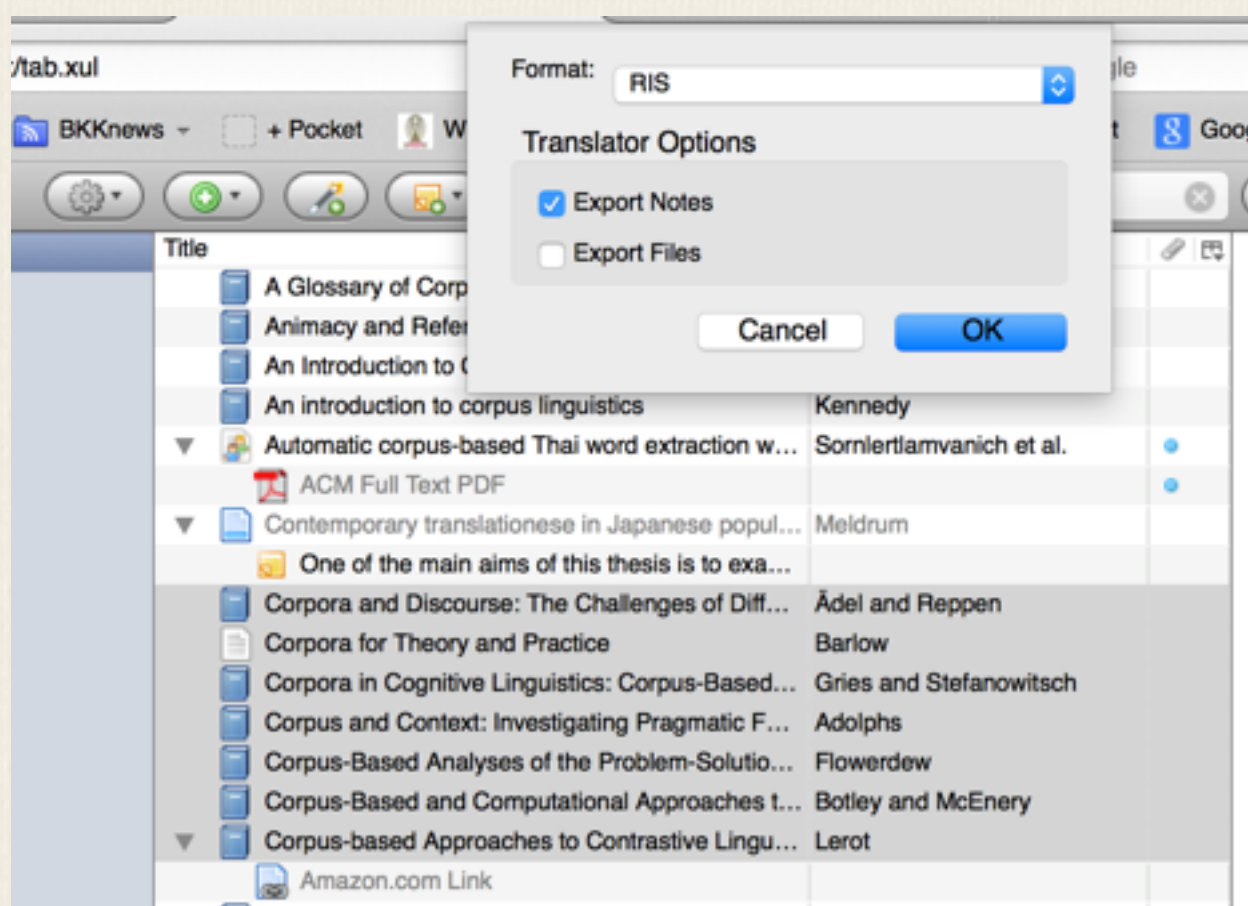
ตัวอย่างสาธิตการใช้งาน Zotero เพื่อเติมรายการข้อมูลต่างๆ ที่พบระหว่างการค้นผ่าน Firefox แสดงในวิดีโอจาก https://www.zotero.org/support/screencast_tutorials การค้นหาข้อมูลภายใน Zotero สามารถพิมพ์คำที่ต้องการค้นในช่อง Search เฉพาะรายการที่ตรงกับคำค้นจะปรากฏให้เห็น

การย้ายข้อมูล Zotero ไป EndNote

การย้ายข้อมูล ZOTERO ไป ENDNOTE

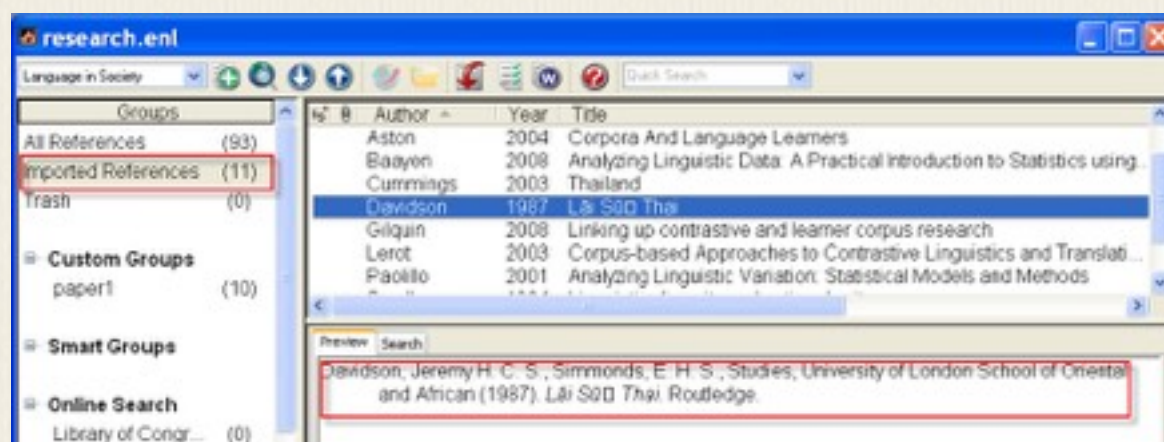
การย้ายข้อมูลบรรณานุกรมจาก Zotero ไปยัง EndNote สามารถทำได้โดยการ export เป็น RIS format แล้วไป import ด้วย EndNote

เลือก Action – Export Library แล้วเลือก format เป็น RIS



เมื่อ save ไฟล์ RIS แล้ว ให้เปิดไฟล์นั้น ข้อมูลต่างๆจะถูกนำเข้าในฐานข้อมูลของโปรแกรม EndNote ที่กำลังใช้งานอยู่ทันที

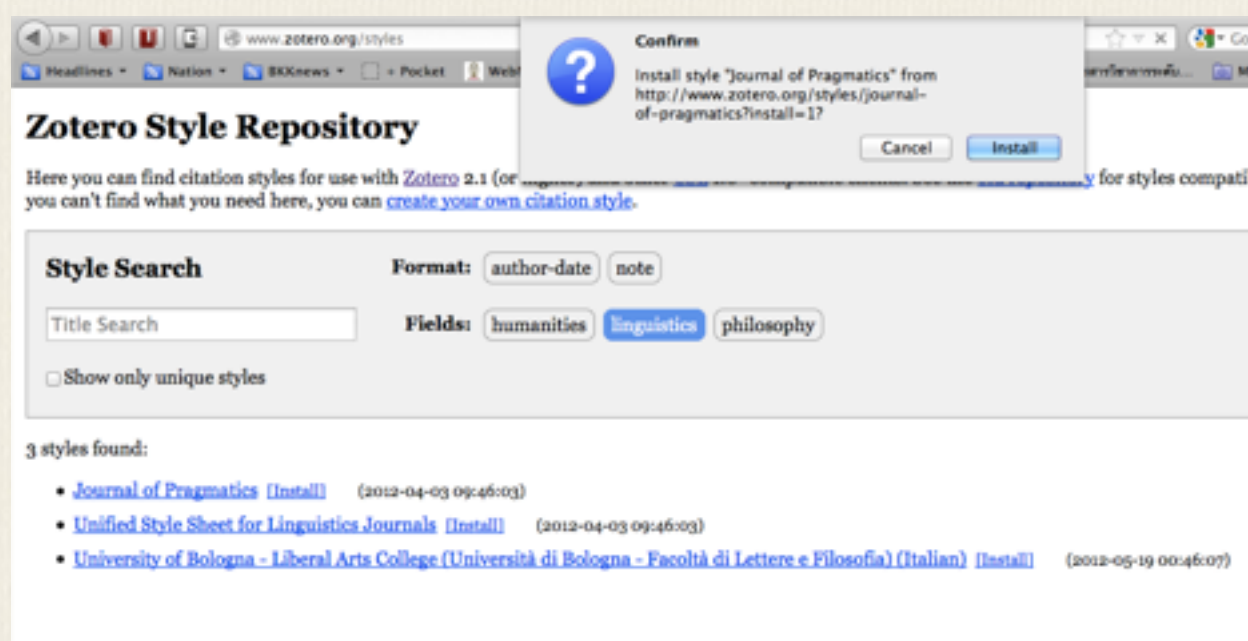
ข้อมูลที่ import มาจะแสดงแยกให้เห็น ข้อมูลหนังสือ หรือแม้แต่เว็บเพจ ก็จะถูกเก็บเข้าใน EndNote และนำไปใช้งานต่อได้



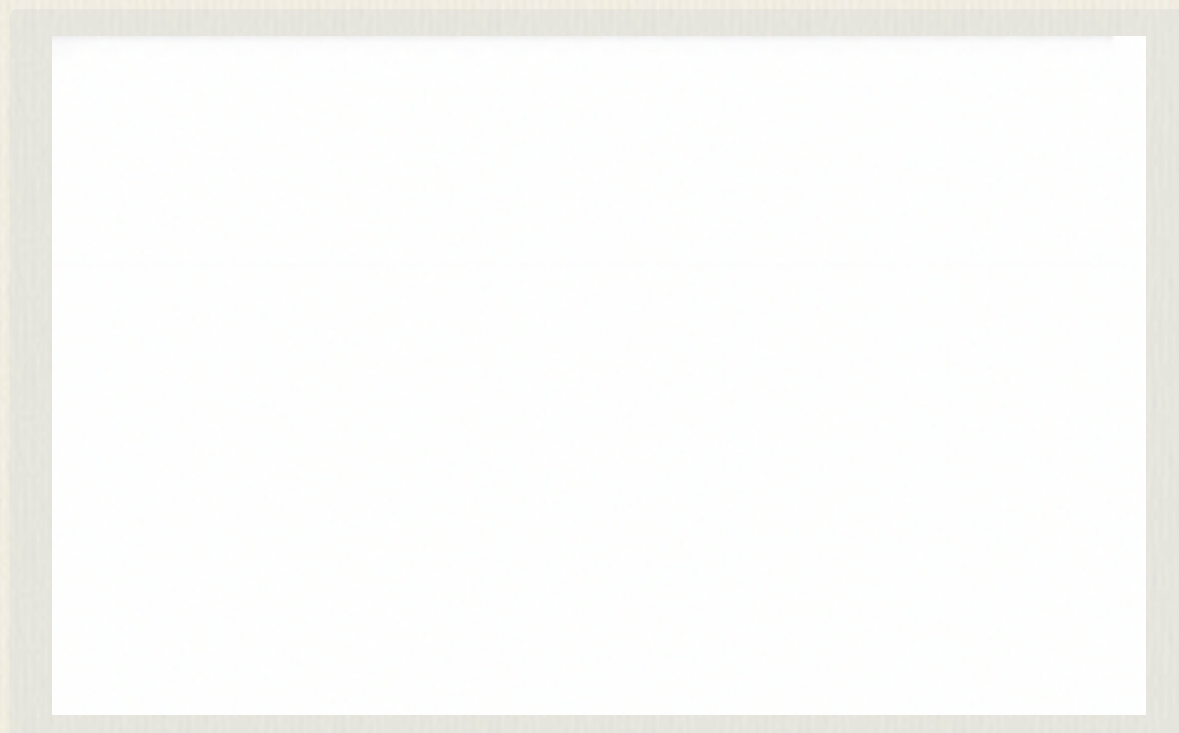
การใช้ Zotero ทำบรรณานุกรม

การใช้ ZOTERO ทำบรรณานุกรม

ความจริง ถ้าไม่มี EndNote จะใช้ Zotero เพื่อช่วยทำบรรณานุกรม และสร้างรายการตาม style ต่างๆ ก็ได้เช่นกัน สามารถติดตั้ง style ที่ไม่มีมาในตอนแรกได้โดยไปที่ <http://www.zotero.org/styles> แล้วเลือก install style ที่ต้องการใช้



โปรแกรม Zotero สามารถทำงานร่วมกับ MS Word ได้ สามารถดูวิดีโอสาธิตได้ที่ http://www.zotero.org/support/word_processor_integration แต่จะต้องติดตั้ง plug-in ของ Word ก่อน ซึ่งสามารถ download ได้ที่ http://www.zotero.org/support/word_processor_plugin_installation



การ backup และ sync Zotero

การ BACKUP และ SYNC ZOTERO

ข้อมูลที่เก็บไว้ใน Zotero จะถูกเก็บโดย default ที่ต่อไปนี้

On a Mac:

/Users/<username>/Library/Application Support/Firefox/Profiles/<randomstring>/zotero

On Windows 2000/XP:

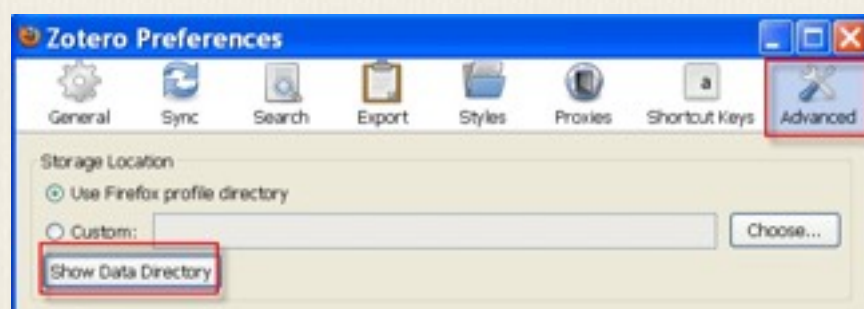
C:\Documents and Settings\<username>\Application Data\Mozilla\Firefox\Profiles\<randomstring>\zotero

On Windows Vista:

C:\Users\<User Name>\AppData\Roaming\Mozilla\Firefox\Profiles\<randomstring>\zotero

ถ้าต้องการ backup ข้อมูล Zotero ไว้ ให้ไปที่ folder ข้างบนแล้ว copy ทั้ง folder เก็บไว้ หากมีปัญหาภายหลังก็นำ folder ที่ copy ไว้มาทบทวนไปได้

สามารถเข้าไปที่ Preference ของ Zotero แล้วเรียกให้แสดง folder ที่เก็บข้อมูลหรือจะบอกให้นำข้อมูลไปเก็บที่ใหม่ที่กำหนดก็ได้ โดยเลือกที่ Custom แล้วระบุ folder ที่ต้องการ



หากใช้ zotero ในหลายๆเครื่อง สามารถ sync ข้อมูลผ่านทาง zotero server ได้ โดยจะต้องมี account บน zotero ก่อน ให้เข้าไปที่ zotero preference เพื่อสร้าง account การ sync ข้อมูลสามารถกดไอคอน sync ที่มุมขวาบนได้

รายละเอียดการใช้งาน Zotero อ่านได้จาก <http://www.zotero.org/support/>



โปรแกรมเพื่องาน ภาษาศาสตร์

Chapter 3

โปรแกรมเพื่องานภาษาศาสตร์

ส่วนนี้จะกล่าวถึง โปรแกรมต่างๆ ที่สามารถนำมาใช้เพื่องานวิจัยทางภาษาศาสตร์ได้

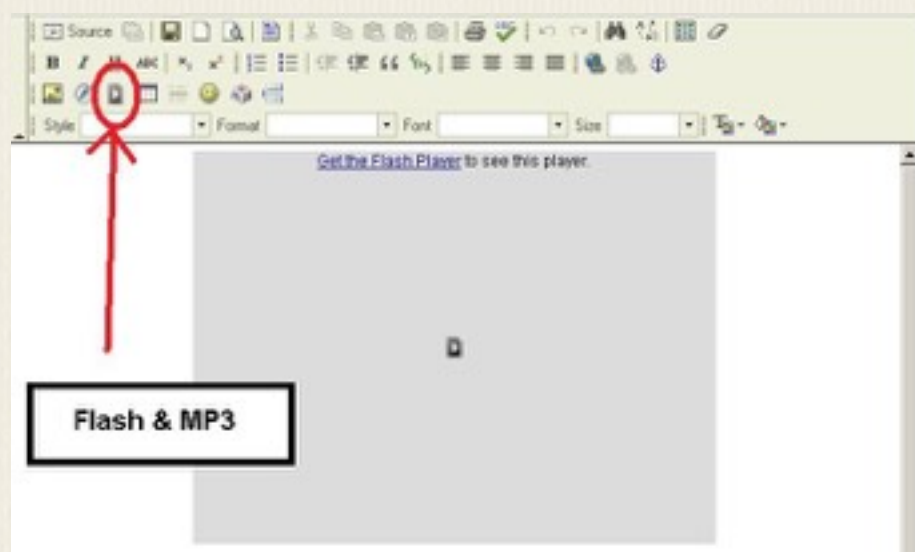
การทำแบบสอบถามผ่านคอมพิวเตอร์

การทำแบบสอบถามผ่านคอมพิวเตอร์

มีเครื่องมือและเว็บไซต์จำนวนมากที่เปิดให้บริการสร้างแบบสอบถามและจัดส่ง link ทางอีเมลให้ทำแบบสอบถาม และสรุปผลให้ ส่วนมากเป็นบริการที่คิดเงิน มีให้ใช้ฟรีได้บ้างแต่ก็มักจะมีข้อจำกัด เช่น จำนวนข้อคำถามไม่เกิน 20 ข้อ จำนวนคนที่ตอบแบบสอบถามไม่เกิน 200 คน ตัวอย่างเช่น <http://www.survs.com/> ใช้บริการฟรีได้ไม่เกิน 200 response แต่บางที่ยอมให้ใช้ได้เป็นปริมาณมาก เช่น <http://www.kwiksurveys.com/> แต่จะคิดค่าบริการจากการให้คำปรึกษาแทน

Kwik surveys <http://www.kwiksurveys.com/>

สามารถแสดงคำถามเป็น text หรือเป็นเสียง mp3 หรือ flash วิดีโอได้



สามารถเลือกสร้างรูปแบบคำตอบได้หลายแบบ ตัวอย่างเช่น

Which of the following do you read every week ?

- ☐ Fiction books
- ☐ Non fiction books
- ☐ Magazines
- ☐ Instruction manuals
- ☐ Newspapers

Please rate the following aspects of our service

	Excellent	Good	Neutral	Fair	Poor	N/A
Room Service	☐	☐	☐	☐	☐	☐
Restaurant	☐	☐	☐	☐	☐	☐
Laundry service	☐	☐	☐	☐	☐	☐
Booking process	☐	☐	☐	☐	☐	☐

Please Enter Your Contact Details

First Name

Last Name

Telephone Number

Email Address

Rank the following items of importance when going to the beach:

Scale:

- 1 Must Have
- 5 Do Not Need.

	1	2	3	4	5
Ice Cream	☐	☐	☐	☐	☐
Sun Cream	☐	☐	☐	☐	☐
Deck Chair	☐	☐	☐	☐	☐
Towel	☐	☐	☐	☐	☐
Sun Hat	☐	☐	☐	☐	☐

Google doc เป็นอีกวิธีหนึ่งที่สามารถใช้สร้างแบบสอบถาม online ได้ฟรี เพียงแค่มี Google account แล้วสร้าง form ขึ้นมาใน Google doc รายละเอียดสามารถอ่านได้จาก tutorial ต่อไปนี้

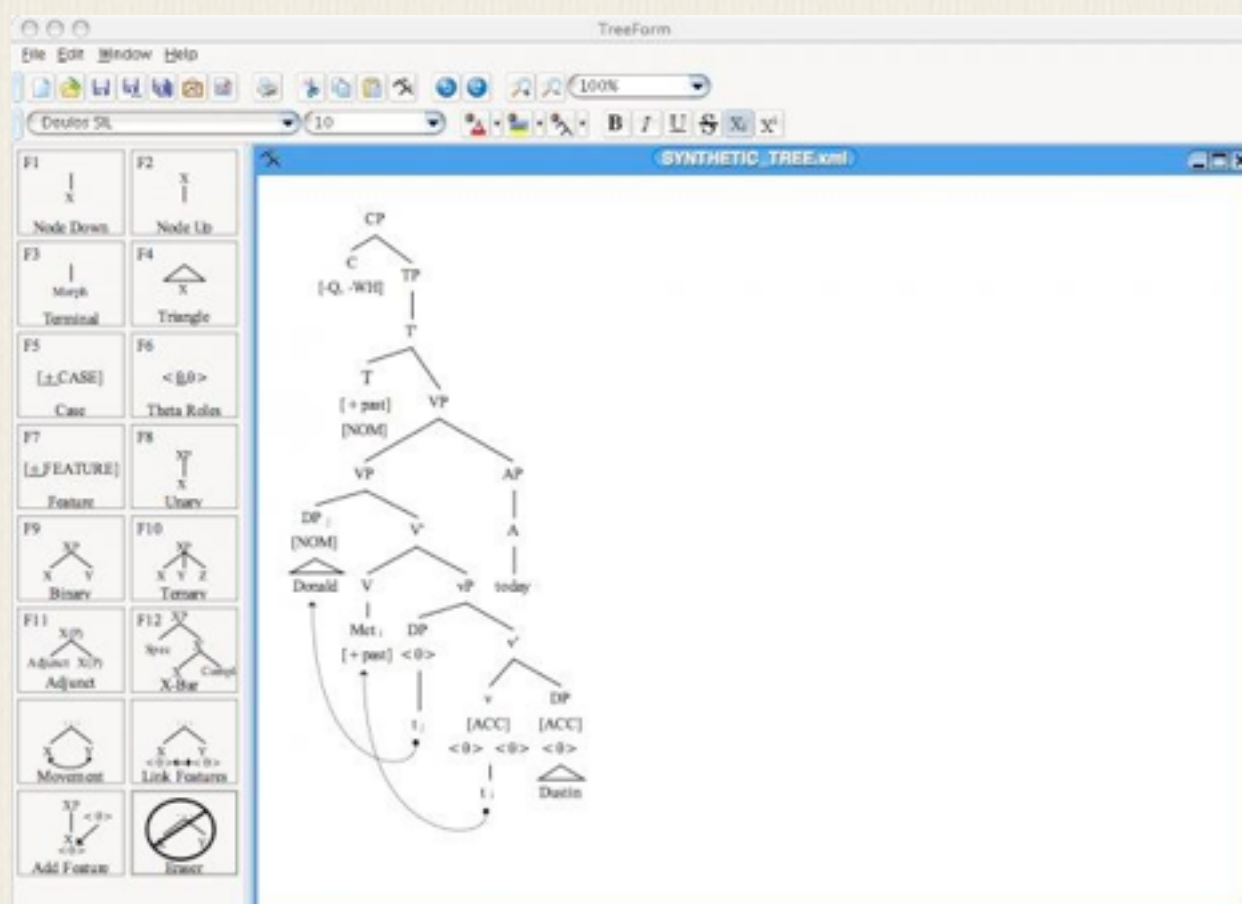
<https://support.google.com/docs/answer/87809?hl=en>

<http://teacherlink.ed.usu.edu/tlresources/training2/Google/GoogleForms.pdf>

<https://www.surveymonkey.com/> เป็นอีกหนึ่งทางเลือกที่สามารถสร้าง online questionnaire ได้ สามารถกำหนดให้หนึ่ง device มีสิทธิการทำหนึ่งครั้ง Google doc ความจริงก็ทำได้ แต่ใช้วิธีการจำกัดการใช้ด้วยการเช็ค account ผู้ที่จะทำแบบสอบถามที่กำหนดแบบนี้จึงต้องมี google account เพื่อ login

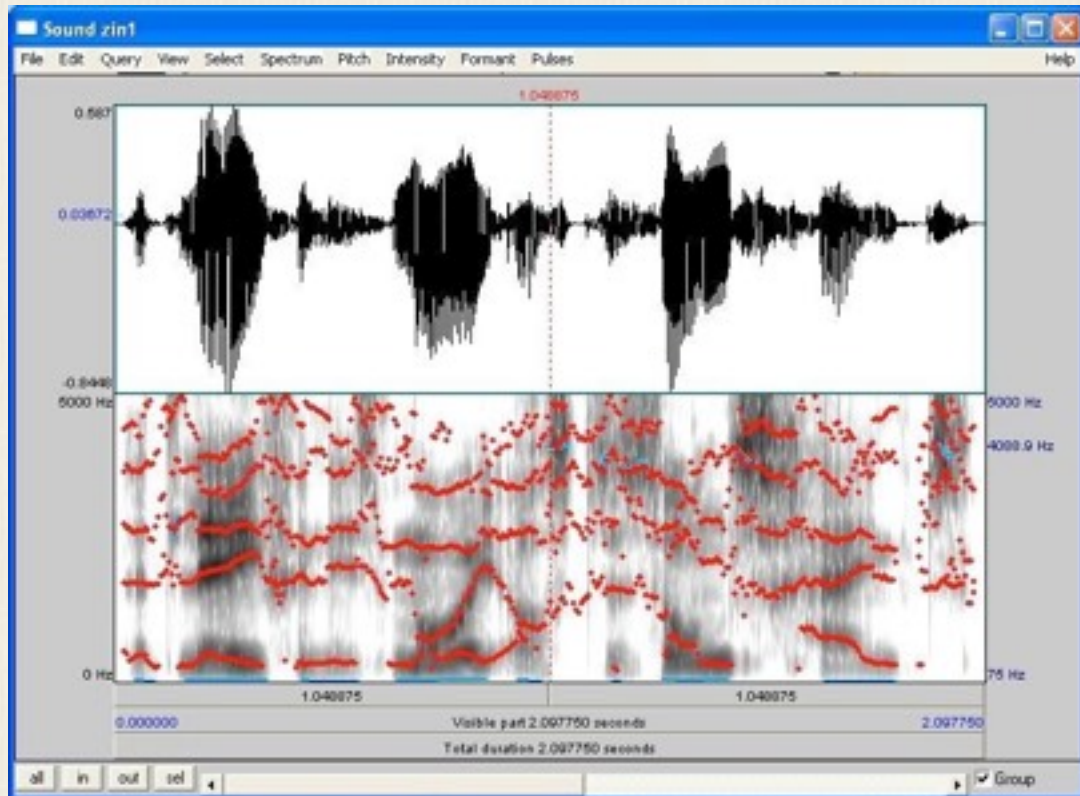
โปรแกรม TreeForm Syntax Tree Drawing

โปรแกรม TREEFORM SYNTAX TREE DRAWING เป็นโปรแกรม OPEN SOURCE สามารถโหลดได้จาก [HTTP://SOURCEFORGE.NET/PROJECTS/TREEFORM/](http://SOURCEFORGE.NET/PROJECTS/TREEFORM/) ด้วยภาษาจาวาจึงใช้กับเครื่อง WINDOWS หรือ MAC ก็ได้ เหมาะสำหรับงานที่จำเป็นต้องเขียนโครงสร้างต้นไม้ โปรแกรมจะช่วยอำนวยความสะดวกในการสร้าง NODE และ BRANCH ต่างๆ ดังภาพตัวอย่างจากโปรแกรมนี้



โปรแกรม Praat

โปรแกรม PRAAT เป็นโปรแกรมที่นิยมใช้ในงานวิจัยทางด้านเสียง สามารถโหลดมาใช้ได้ฟรีจาก [HTTP://WWW.FON.HUM.UVA.NL/PRAAT/](http://www.fon.hum.uva.nl/PRAAT/) มีทั้งที่ทำงานบน WINDOWS และ MAC สามารถใช้ในงาน SPEECH ANALYSIS, SPEECH SYNTHESIS, ทำ LABEL AND SEGMENTATION, MANIPULATE SPEECH เป็นต้น ผู้สนใจสามารถอ่านคำแนะนำการใช้งานโปรแกรม PRAAT เพื่องานทางภาษาศาสตร์ได้ที่ [HTTP://SAVETHEVOWELS.ORG/PRAAT/](http://savethevowels.org/PRAAT/) หรืออ่านคู่มือเบื้องต้นได้ที่ [HTTP://WWW.STANFORD.EDU/DEPT/LINGUISTICS/CORPORA/MATERIAL/PRAAT_WORSHOP_MANUAL_v421.PDF](http://www.stanford.edu/dept/Linguistics/corpora/material/PRAAT_WORSHOP_MANUAL_v421.pdf) และ [HTTP://SAVETHEVOWELS.ORG/PRAAT/UsingPRAATforLinguisticResearchLatest.PDF](http://savethevowels.org/PRAAT/UsingPRAATforLinguisticResearchLatest.pdf)



โปรแกรมสร้างการทดลอง ภาษาศาสตร์จิตวิทยา

โปรแกรมสร้างการทดลองภาษาศาสตร์จิตวิทยามีหลายโปรแกรมที่สามารถใช้ได้

- DMDX (<http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm>)

เป็นฟรีโปรแกรมรุ่นแรกๆ จึงทำงานบน DOS หรือ windows XP นำเสนอคำถามที่เป็นภาพและเสียง แล้วเก็บ response กับ reaction time จากคำตอบผู้ทดลอง ผู้สนใจสามารถดู Tutorial ได้ในเว็บของ Matt Davis

(<http://www.mrc-cbu.cam.ac.uk/people/matt.davis/dmdx.html>)

- E Prime (<https://www.pstnet.com/eprime.cfm>)

เป็นโปรแกรมจำหน่ายสำหรับการทดลองทางภาษาศาสตร์จิตวิทยาที่ได้รับความนิยมเพราะสะดวกในการใช้งาน

- Psyscope X (<http://psy.ck.sissa.it/>)

เป็นโปรแกรมฟรีทำงานบน OS X นำเสนอคำถามที่เป็นภาพและเสียง แล้วเก็บ response กับ reaction time จากคำตอบผู้ทดลองได้ ปัจจุบันกำลังปรับปรุงให้มี GUI ที่ง่ายขึ้นต่อการออกแบบการทดลอง แต่ไม่สามารถใช้งานบน Windows

- PXLab (<http://irtel.uni-mannheim.de/pxlab/index.html>)

เป็นอีกโปรแกรมสำหรับงานทดลองทางจิตวิทยา เขียนด้วยภาษา Java เพื่อทำการทดลองรูปแบบต่างๆ ได้

- Psychopy (<http://psychopy.org/>)

เป็นโปรแกรมฟรีเขียนด้วยภาษา Python สามารถใช้ได้ทั้งกับ Windows และ Max OS X มี GUI ที่ช่วยในการออกแบบการทดลองและกำหนดค่าต่างๆ ที่ต้องการใช้ในการทดลอง เป็นตัวเลือกที่สามารถใช้แทน E Prime ได้ Tutorial video ที่เห็นจะช่วยให้เข้าใจการใช้ Psychopy ในการออกแบบและสร้างการทดลองง่าย ๆ ที่แสดงข้อความและรอรับคำตอบจากผู้ทดลอง



Build your first PsychoPy experiment (Stroop task) <https://youtu.be/VV6qhuQgsil>

<http://wordij.net/>

โปรแกรมดึงข้อมูลจากเว็บ

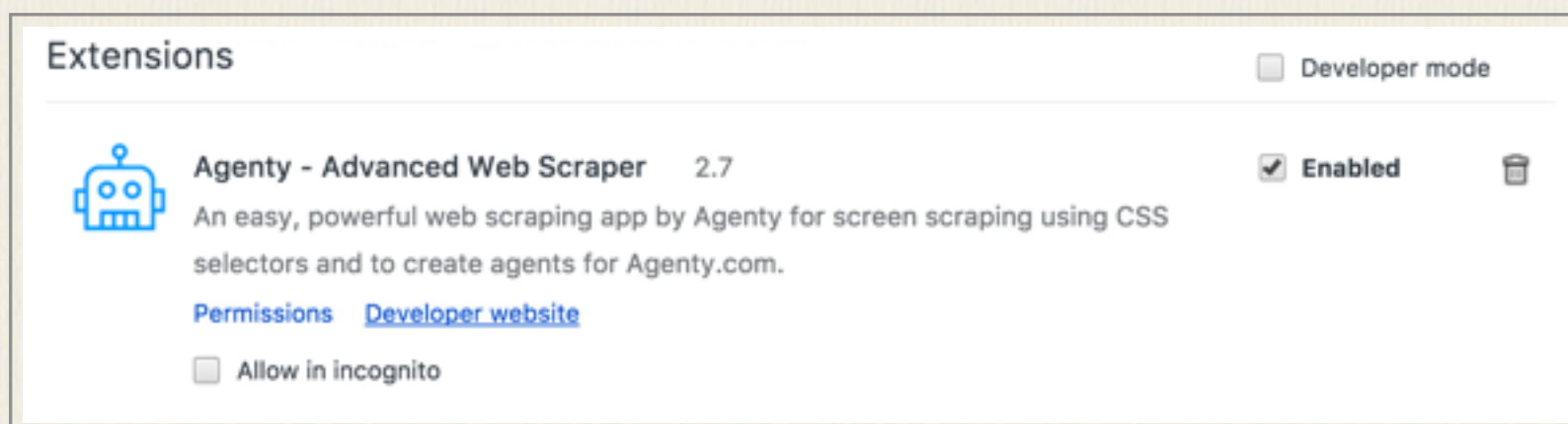
โปรแกรม Httrack

(<https://www.httrack.com/>) เป็นโปรแกรม copy website เหมาะสำหรับการคัดลอกข้อมูลบนเว็บมาเก็บไว้ในเครื่องเพื่อใช้งาน offline Httrack เป็นโปรแกรมสำหรับติดตั้งบน Windows สะดวกในการเก็บข้อมูลจากเว็บใดเว็บหนึ่ง

ใช้บริการ boilerpipe

(boilerpipe-web.appspot.com/) เป็นบริการ web api ที่แสดงการใช้งาน boilerpipe Java library ในการดึงตัวบทจากเว็บที่ระบุ สามารถใช้ดึงข้อมูลตัวบทออกมาเป็น plain text หรือ json ก็ได้ มี option ของการสกัดตัวบทให้เลือกตามที่กำหนด

ใช้ Chrome Extension

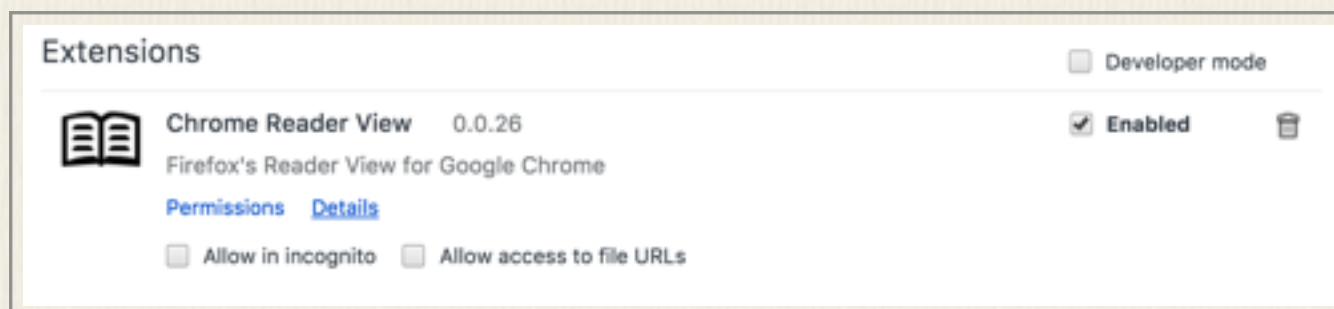


ติดตั้ง Advanced Web Scraper เมื่อพบเว็บที่มีข้อมูลที่ต้องการ ให้กดเรียก extension นี้ แล้วเลือกข้อมูลที่ต้องการโดยการกด * แล้วคลิกตำแหน่งข้อมูลที่ต้องการบนหน้าเว็บนั้น จะเห็นแถบสีเขียวข้อความที่จะถูกสกัดออกมา จากนั้นให้ตั้งชื่อ field ที่แทนข้อมูลนั้น ทำเช่นนี้กับข้อมูลอื่นจนครบทุก field ข้อมูลที่ต้องการ จากนั้นกด preview เพื่อดูข้อมูลที่จะ extract แล้วเลือก option ว่าจะบันทึกข้อมูลเป็น csv, json โปรแกรมนี้เหมาะสำหรับการดึงเว็บที่มีข้อมูลหลายรายการที่ดึงออกมาเป็นแต่ละ record ได้ เช่น

หน้าเว็บวิจารณ์เรื่องต่างๆ ที่สามารถดึง review แต่ละคนออกมาเป็น 1 รายการได้ แต่ข้อมูลที่ดึงจะยังคงติด html tag อยู่บ้าง ต้องมาหาทางตัดออกเองภายหลังหากไม่ต้องการ



ใช้ Reader View addon เพื่อแปลงหน้าเว็บให้อยู่ในรูป text based เป็นหลักก่อน addon นี้ใช้เพื่อให้สะดวกต่อการอ่าน ซึ่งก็ทำให้สะดวกต่อการ copy-paste เฉพาะข้อความที่ต้องการออกมาได้ วิธีนี้เป็นการ copy หน้าเว็บเองเป็นหลัก



ใช้ R package สำหรับ web crawler

เราสามารถ ใช้ package ใน R เพื่อช่วยในการสกัดข้อมูลที่ต้องการจากเว็บเพื่อมาใช้ในงานที่ต้องการได้ package หนึ่งที่เป็นประโยชน์เพื่อการนี้คือ rvest ตัวอย่างข้างล่าง ดัดแปลงจาก stat4701.github.io/edav/2015/04/02/rvest_tutorial/

```
> install.packages("rvest")
```

```
> library(rvest)
```

load ข้อมูลจากหน้าเว็บ review ภาพยนตร์ Arrival มาเก็บไว้

```
> arrival_movie <- html("http://www.imdb.com/title/tt2543164/")
```

ดึงข้อมูล rating ออกมาเก็บไว้ที่ตัวแปร rating ข้อมูลนี้อยู่ในแท็ก strong ซึ่งภายใต้มีแท็ก span อยู่แล้ว สกัด text ในนั้นออกมา แปลงเป็นตัวเลข เครื่องหมาย %>% เป็นการ pipe หรือส่งผ่านข้อมูลต่อ ซึ่งจะเห็นว่าในคำสั่งให้เอาข้อมูลเว็บที่เก็บใน arrival_movie มาสกัด node ที่มีแท็ก strong แล้วก็ span เอาข้อมูลที่ได้ส่งต่อให้ html_text() ดึง text ออกมาส่งต่อให้มาแปลงเป็นตัวเลข

```
> rating <- arrival_movie %>%
```

```
+ html_nodes("strong span") %>%
```

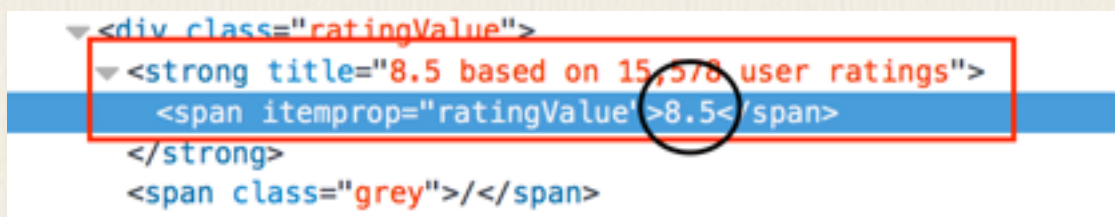
```
+ html_text() %>%
```

```
+ as.numeric()
```

```
> rating
```

หากดูภายในเว็บจะเห็นข้อมูลการให้คะแนนอยู่ในแท็กนี้

```
<strong title="8.5 based on 15,578 user ratings"><span  
itemprop="ratingValue">8.5</span></strong>
```

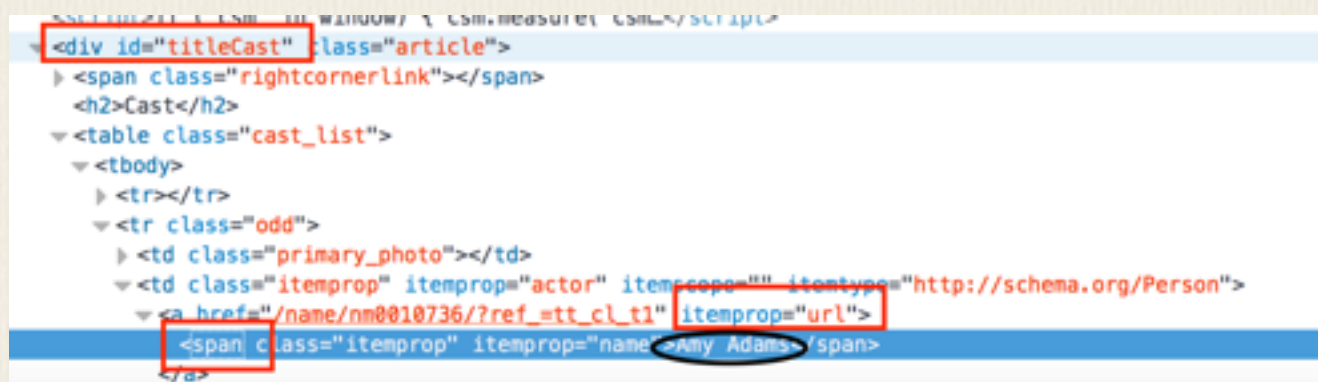


คำสั่งต่อมาสั่งให้หาโหนดที่มี attribute value เป็น titleCast (มีเครื่องหมาย # นำหน้า)

แล้วดูต่อหา attribute ชื่อ itemprop (มีเครื่องหมาย . นำหน้า) แล้วต่อไปจนพบแท็ก span

จึงส่งข้อมูลต่อให้ดึง text ออกมาซึ่งจะเป็นชื่อนักแสดง และเนื่องจากพบข้อมูลแบบนี้มากกว่าหนึ่ง จึงดึงทั้งหมดออกมาเป็นรายการชื่อทั้งหมดได้

```
> cast <- arrival_movie %>%  
+ html_nodes("#titleCast.itemprop span") %>%  
+ html_text()  
> cast  
[1] "Amy Adams"      "Jeremy Renner"   "Michael Stuhlbarg" "Forest Whitaker"  
"Sangita Patel"  
[6] "Mark O'Brien"   "Abigail Pniowsky" "Tzi Ma"           "Nathaly Thibault" "Ruth  
Chiang"  
[11] "Jadyn Malone"    "Julia Scarlett Dan" "Russell Yuen"      "Anana Rydvald"  
"Leisa Reid"
```



จะเห็นว่าการใช้ rvest ช่วยให้เราสามารถดึงข้อมูลส่วนที่ต้องการในหน้าเว็บมาใช้ได้เลย เพียงแต่เราต้องรู้ว่าข้อมูลนั้นอยู่ที่ไหน ภายใต้อะไร เพื่อจะได้เขียนคำสั่งสกัดให้ถูกตำแหน่งได้

- ใช้ Python package สำหรับ web crawler

ใช้โปรแกรม Octoparse

Octoparse เป็น โปรแกรมบน Windows มี version ฟรีให้ใช้

โปรแกรม Corpus Annotator

โปรแกรม ANNOTATOR คือ โปรแกรมที่ช่วยในการ MARKUP ข้อมูลเพิ่มเติมใน ตัวย่อ โปรแกรมบางตัวออกแบบมาเพื่อใช้เหมือนเป็นการทำ NOTE เพิ่มเติมใน เอกสาร เช่น โปรแกรม ANNOTATOR ([HTTP://ANNOTATORJS.ORG](http://annotatorjs.org)) ที่เขียนด้วย ภาษา JAVASCRIPT ให้ผู้ใช้สามารถเติมข้อความเพิ่มเติมไปในเอกสารผ่านทางเว็บ ได้ ไม่ว่าจะเป็นหน้าเว็บที่เป็น HTML หรือ PDF หรือ EPUB แต่ก็ไม่เหมาะกับการ กากับข้อมูลทางภาษาศาสตร์เพื่อสร้างข้อมูลที่จะนำมาวิเคราะห์ต่อ

โปรแกรม BRAT RAPID ANNOTATION TOOL ([HTTP://BRAT.NLPLAB.ORG](http://brat.nlplab.org)) เป็นโปรแกรมที่ทำงานบนเว็บสำหรับกากับข้อมูลทางภาษา เช่น ชื่อเฉพาะ หมวดคำ ความสัมพันธ์ระหว่างคำ จึงเป็นโปรแกรมที่เหมาะกับการกากับข้อมูลภาษาศาสตร์ เพื่อนำมาวิเคราะห์ต่อมากกว่า

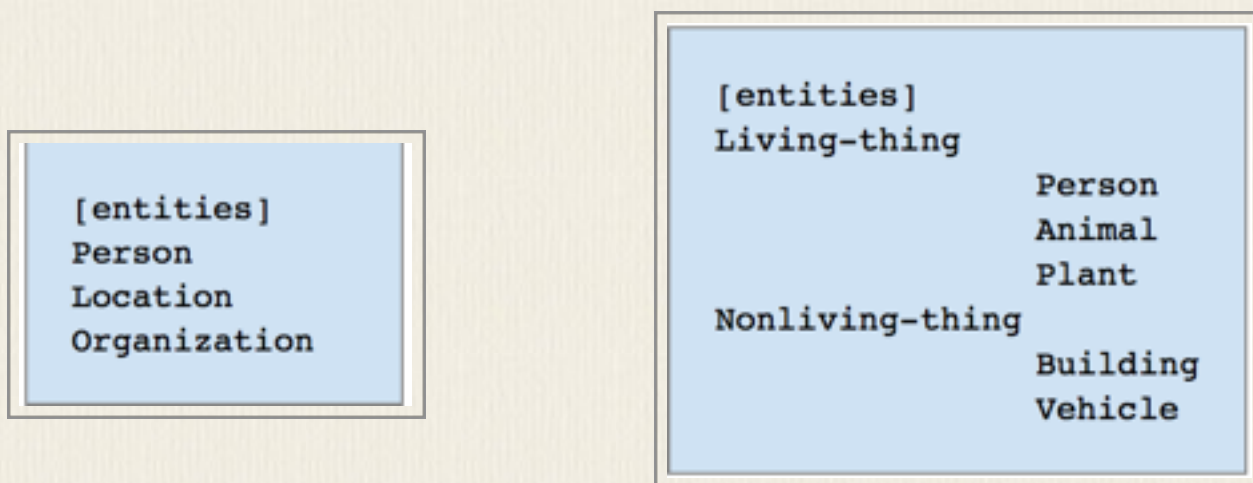
โปรแกรมนี้ต้องติดตั้งบน SERVER ทำงานด้วยโปรแกรม PYTHON 2.5 ขึ้นไป ขั้นตอนนี้จะต้องให้ผู้ดูแล SERVER เป็นคนติดตั้งให้ก่อนจึงจะใช้งานได้ เมื่อติดตั้ง เรียบร้อย ผู้ที่ต้องการใช้ จะต้องมีไฟล์ข้อมูล และไฟล์ CONFIGURATION ที่กำหนด แท็กต่างๆ และความสัมพันธ์ที่จะมีระหว่างแท็กได้ แล้ว UPLOAD ไฟล์เหล่านี้ขึ้นไป ที่ SERVER เพื่อใช้งานได้

ไฟล์กำหนด THE CONFIGURATION ของการทำ ANNOTATION มีสี่ไฟล์ คือ

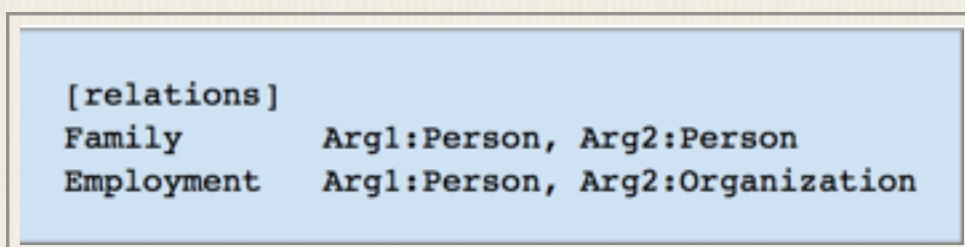
- ANNOTATION.CONF: ANNOTATION TYPE CONFIGURATION
- VISUAL.CONF: ANNOTATION DISPLAY CONFIGURATION
- TOOLS.CONF: ANNOTATION TOOL CONFIGURATION
- KB_SHORTCUTS.CONF: KEYBOARD SHORTCUT TOOL CONFIGURATION

ไฟล์ ANNOTATION.CONF ประกอบด้วยสี่ส่วน คือ [ENTITIES] [RELATIONS] [EVENTS] และ [ATTRIBUTES] [ENTITIES] ใช้กำหนดแท็กสำหรับกากับแต่ละหน่วย เช่น PERSON, ORGANIZATION, LOCATION หน่วยที่จะมีทั้งหมดปรากฏล่าง [ENTI-

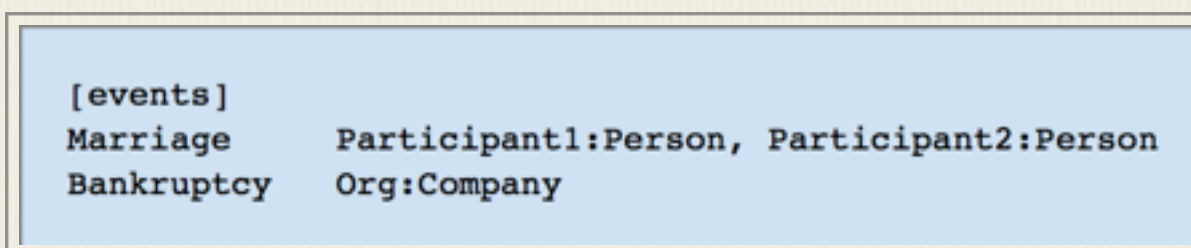
TIES] หนึ่งบรรทัดต่อหนึ่งหน่วย ENTITIES อาจมีลำดับชั้นได้ ด้วยการจัดเรียงแล้ว
เคาะแท็บหน้ากลุ่มเป็นลูกกลุ่ม



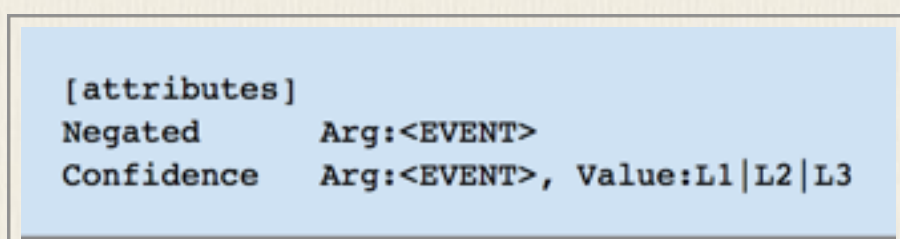
[RELATIONS] ใช้กำหนดความสัมพันธ์ระหว่าง 2 ENTITY เช่น กำหนด ARGUMENT1 เป็น PERSON ARGUMENT2 เป็น ORGANIZATION สำหรับความสัมพันธ์ EMPLOYMENT ตามตัวอย่างนี้



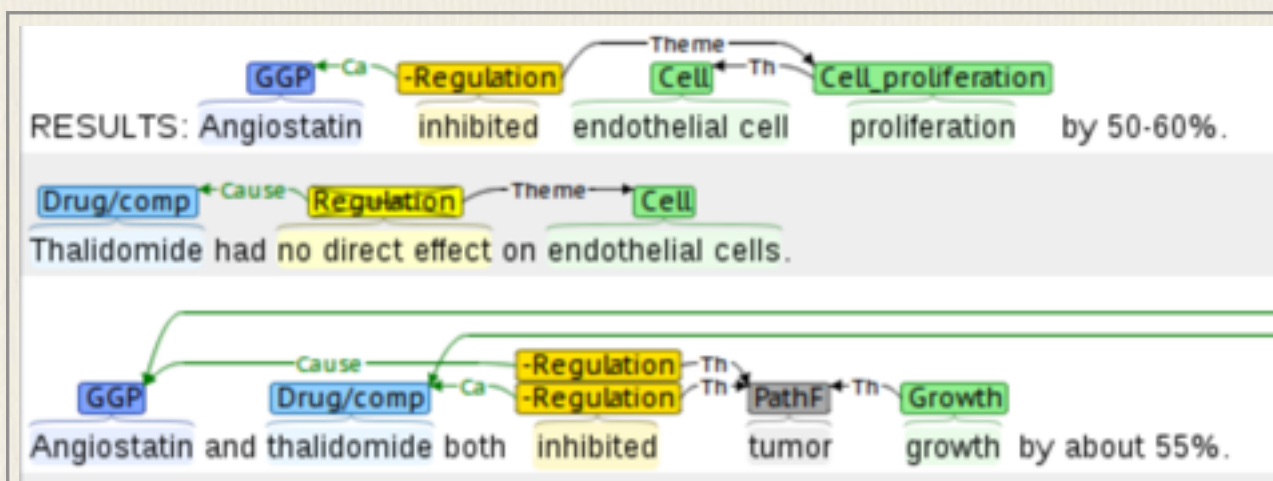
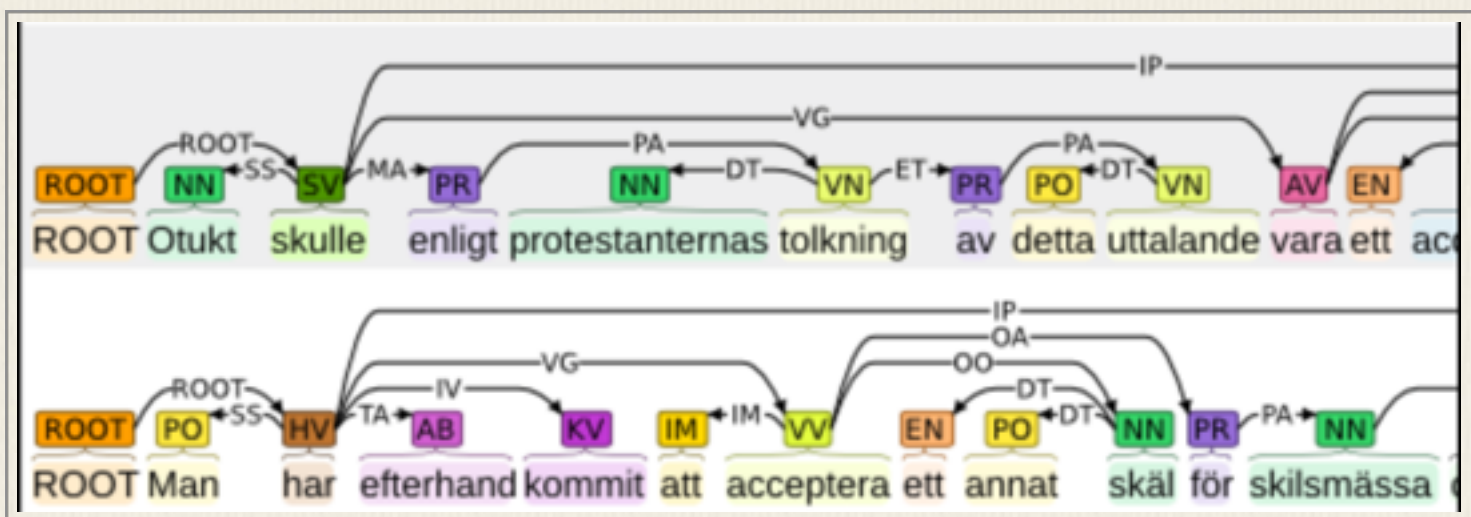
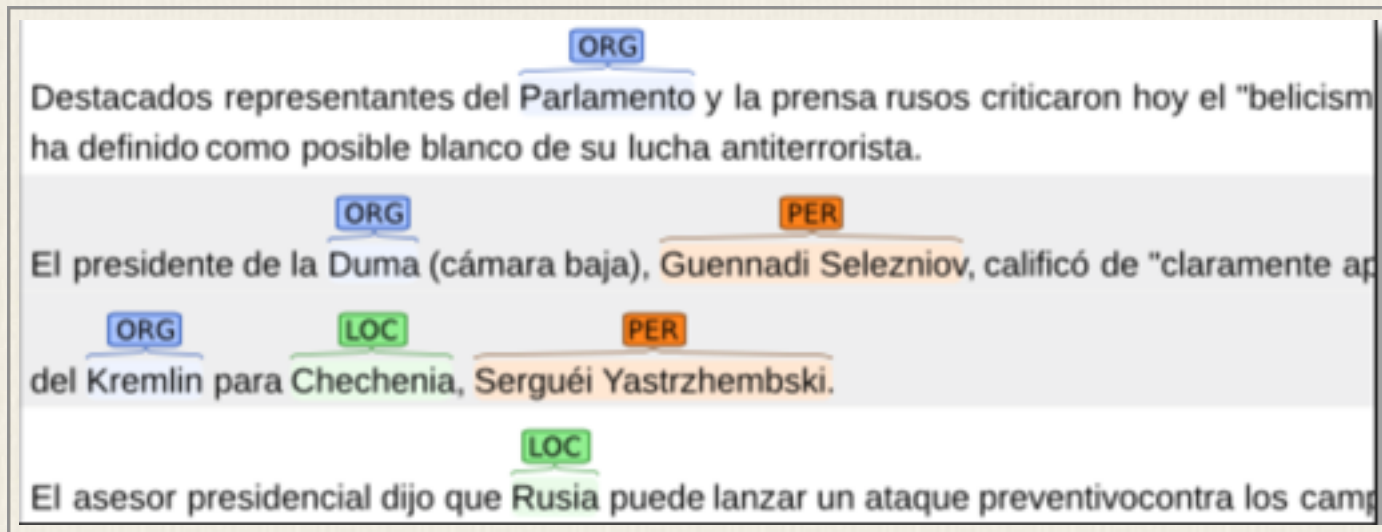
[EVENTS] ใช้กำหนดความสัมพันธ์ของ ENTITY หรือ RELATION ที่มีใน EVENT นั้น จะมีองค์ประกอบหนึ่งตัวหรือมากกว่าสองก็ได้ มีลักษณะคล้ายการกำหนด FRAME ของเหตุการณ์ว่าประกอบด้วยอะไรบ้าง องค์ประกอบเขียนในรูป ROLE:TYPE เช่น PARTICIPANT1:PERSON ORG:COMPANY เช่นตัวอย่างนี้



[ATTRIBUTES] ใช้กำหนด VALUE ให้กับ ANNOTATION หรือแท็กต่างๆ ที่กำกับ จะมีค่าเป็น BINARY หรือมีค่าเป็น ตัวเลือกต่างๆที่กำหนดก็ได้



สำหรับ CONFIGURATION อื่นเป็นเรื่องการกำหนดการแสดงผลหน้าจอและสี
 ต่างๆ ที่แสดง รายละเอียดสามารถดูได้จากเว็บ โครงการเอง BRAT ANNOTATION นี้
 ได้ถูกใช้สำหรับกำกับข้อมูลทางภาษาศาสตร์หลายเรื่อง เช่น การกำกับชื่อเฉพาะ
 การกำกับโครงสร้างต้นไม้แบบพืงพา การกำกับ EVENT ดังตัวอย่างที่แสดงข้างล่าง
 นี้



โปรแกรมสร้างไฟล์ pdf copy

โปรแกรมสร้างไฟล์ PDF

เราสามารถพิมพ์บทความหรือวิทยานิพนธ์เป็นไฟล์ pdf ได้ หากใช้ Mac จะเลือกได้จากคำสั่ง Print แล้วเลือก Save as PDF หากใช้ Windows สามารถติดตั้งโปรแกรม CutePDF ซึ่งจะช่วยให้มีตัวเลือกเครื่องพิมพ์เพิ่มอีกหนึ่งตัวคือ CutePDF Writer โปรแกรม CutePDF สามารถโหลดได้จากเว็บ (<http://www.cutepdf.com/>)



การดูแลรักษาข้อมูล

Chapter 4

การดูแลรักษาข้อมูล

ข้อมูลงานวิจัยไม่ว่าจะเป็นข้อมูลดิบ ข้อมูลที่วิเคราะห์แล้ว ข้อสรุปหรือบทความ
วิทยานิพนธ์ รายงานวิจัย ถือเป็นสิ่งสำคัญที่จะต้องปกป้องรักษาตลอดช่วงของการทำ
วิจัย เพราะหากมีการสูญหายของข้อมูลไม่ว่าจะในขั้นตอนใด ก็อาจทำให้งานวิจัยนั้น
ล้มเหลวไม่สำเร็จตามเวลาที่กำหนด

การสำรองข้อมูลและการรักษาความปลอดภัย

การสำรองข้อมูลและการรักษาความปลอดภัย

การสำรองข้อมูล เป็นกิจวัตรที่ผู้วิจัยควรต้องกระทำอย่างสม่ำเสมอทุกวัน หรืออย่างน้อยทุกสัปดาห์ ข้อมูลควรจะต้องมีการสำรองไว้หลายชุด และเก็บไว้ในหลายที่ทั้งที่บ้าน ที่ทำงานหรือมหาวิทยาลัย หรือที่เป็น cloud storage เช่น dropbox, google drive หรือ one-drive หรือบางครั้งอาจสำรองข้อมูลเป็นหลาย ๆ version คือไม่ได้ลบไฟล์เก่าทิ้ง แต่เก็บไฟล์ข้อมูลใหม่ วิธีนี้จะทำให้มีข้อมูลเก่าเมื่อสัปดาห์ก่อน เมื่อเดือนก่อน อยู่ตลอดเวลาการวิจัย ทำให้หากต้องการย้อนกลับไป ณ เวลาเดิมก็สามารถทำได้ หากจะทำเช่นนี้สิ่งที่จะต้องคำนึงคือการวางระบบให้ชัดเจนว่าข้อมูลชุดไหนเป็นของช่วงเวลาไหน

ในกรณีที่ข้อมูล (ทั้งที่เป็นข้อมูลภาษา ข้อมูลงานวิจัย ฯลฯ) มีจำนวนมาก อาจสะดวกกว่าถ้าใช้วิธีการ sync ข้อมูลแทนที่จะ copy หรือ backup ไฟล์ทั้งหมด โดยสามารถใช้โปรแกรม เช่น Synkron (<http://synkron.sourceforge.net/>) ซึ่งเป็น โปรแกรมฟรี สำหรับทั้ง PC และ Mac Allway Sync (<http://allwaysync.com/>) สำหรับ PC ซึ่งมีทั้งที่ฟรี และต้องซื้อ หรือ โปรแกรมที่ขายอย่าง GoodSync ได้ <http://www.goodsync.com/> แต่ทั้งนี้ต้องรอบคอบและระวังไม่ให้ข้อมูลเก่าทับข้อมูลใหม่ หรือลบข้อมูลใหม่ทิ้ง ควรทดลองใช้โปรแกรมกับข้อมูลทดสอบอื่นๆ ก่อนจนเข้าใจวิธีการใช้โปรแกรมดีแล้วจึงจะใช้กับข้อมูลจริงได้

การ backup ระบบคอมพิวเตอร์ก็จำเป็นกรณีที่เครื่องที่ใช้มีปัญหา จะได้ restore กลับได้โดยง่าย Windows restore เป็นการใช้งานพื้นฐานที่ช่วยให้เราย้อนระบบกลับไปยังวันที่ไม่มีปัญหาได้ แต่ก็อาจจะไม่สำเร็จในบางกรณี Time machine ของ Mac ใช้สำหรับ backup เครื่องเป็นระยะๆ เพื่อสามารถย้อนกลับไปวันก่อนจะมีปัญหาได้

การ clone หรือสร้าง backup image ของ drive ไว้เป็นอีกวิธีที่ช่วยได้ ทำให้สามารถ re-store ทั้ง hard drive กลับมาได้เมื่อไม่สามารถแก้ไขปัญหาเครื่องได้ โปรแกรมประเภทนี้

ได้แก่ Norton Ghost, ที่เป็น โปรแกรมฟรีให้ใช้ก็มี เช่น Macrium Reflect Free Edition สำหรับ PC (<http://www.macrium.com/ReflectFree.asp>), สำหรับ Mac สามารถใช้ โปรแกรม SuperDuper

(<http://www.shirt-pocket.com/SuperDuper/SuperDuperDescription.html>)

อย่างไรก็ตาม ก่อนใช้งาน โปรแกรมเกี่ยวกับการจัดการ partition ให้ทดลองให้แน่ใจว่า ขั้นตอนเป็นอย่างไร ก่อนจะใช้งานกับข้อมูลที่สำคัญ ในการ backup ควรใช้ external hard-disk ในการ backup drive เมื่อมีปัญหา ก็สามารถ restore จาก external harddisk นั้นได้ โปรแกรมประเภทนี้จำเป็นจะต้องให้เราสร้าง CD สำหรับ boot เครื่องขึ้นมาเองเวลาที่ ต้องการ restore ให้สร้าง boot CD แล้วเก็บไว้ในที่ที่ปลอดภัย หรือถ้าสามารถสร้าง backup บน usb harddisk แล้วกำหนดให้เครื่อง boot จาก usb drive ได้เลยก็จะสะดวก

ไวรัสคอมพิวเตอร์

ไวรัสคอมพิวเตอร์

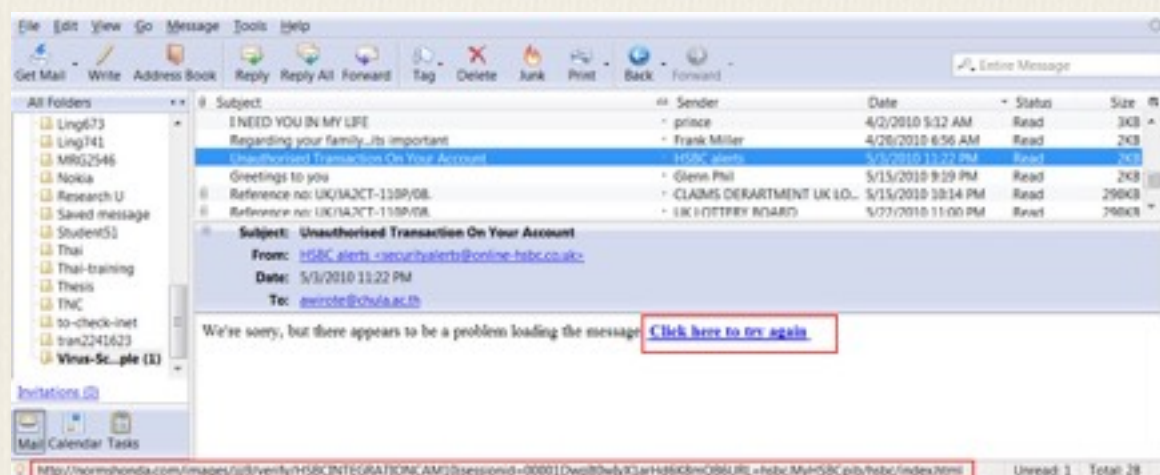
ผู้วิจัยควรมีโปรแกรม Antivirus ติดตั้งและต้องตั้งค่าให้ update ตัวเองอยู่ทุกวัน เครื่องที่ใช้จึงควรต่ออินเทอร์เน็ตอยู่เสมอเพื่อให้สามารถนำข้อมูลใหม่มา update ได้ โปรแกรมป้องกันไวรัสที่ได้แถมมากับเครื่องมักมีระยะเวลาให้ใช้ได้ช่วงหนึ่ง เช่น Norton, McAfee เมื่อพ้นกำหนดแล้วถ้าไม่จ่ายเงิน จะไม่ได้ update ตัวเอง จึงเป็นอันตรายที่ไวรัสใหม่ๆ จะเข้ามาได้ง่ายๆ หากเราไม่รู้และไม่ได้จ่ายเงินรายปีต่อ ทางที่ดีคือให้ใช้ โปรแกรม Anti virus ที่มี free version ให้ใช้ อย่างน้อยหนึ่งตัว เช่น AVG (<http://free.avg.com/>), Avast (<http://www.avast.com/free-antivirus-download>), Avira (<http://www.free-av.com/>) และที่สำคัญ จะต้องคอย update โปรแกรมต่างๆ บนเครื่อง ไม่ว่าจะเป็น windows, Internet Explorer, Firefox, etc. เพราะโปรแกรมเก่าจะมีช่องว่างให้ไวรัสหลุดเข้ามาได้ง่าย

ไวรัสคอมพิวเตอร์เป็น โปรแกรมที่เขียนขึ้นมาเพื่อแฝงตัวในคอมพิวเตอร์และแพร่กระจายไปคอมพิวเตอร์เครื่องอื่นๆ ไวรัสสามารถแพร่ผ่านสื่อต่างๆ diskette, flash drive, CD, internet ผลกระทบที่เกิดจากการติดไวรัสมีความแตกต่างกันไปหลายแต่ไวรัสนั้นถูกเขียนมาเพื่อให้ทำอะไร

โปรแกรมไวรัสมีหลายประเภท มีที่เป็นแบบไวรัสคือต้องอาศัยไฟล์อื่นเป็น host ในการแพร่ ที่เป็นแบบ Trojan คือแฝงมาใน โปรแกรมที่คิดว่าจะนำมาใช้งานหนึ่งแต่แอบซ่อนโปรแกรมร้ายไว้ที่เป็นแบบ worm คือกระจายตัวได้เองพยายามแพร่ตัวเองผ่านเครือข่ายจนระบบล่มไป

ปัจจุบันไวรัสมาในรูปแบบต่างๆ ส่งผ่านอีเมลล์หรือเว็บ เพื่อหลอกผู้ใช้ให้เปิดดู เมื่อเปิดก็จะถูกจู่โจมและติดทันที บางครั้งมาในรูปการหลอกว่าเครื่องเราได้ติดไวรัส หากต้องการ remove หรือป้องกันให้ install โปรแกรมป้องกันที่ให้โดยด่วน วิธีป้องกันคือไม่

เปิดไฟล์ที่ไม่ใช่โปรแกรมที่เรารู้จักและได้ติดตั้งเองในเครื่อง ไม่เปิดไฟล์คนแปลกหน้า หรือแม้แต่ไฟล์ที่ส่งมาจากคนรู้จัก หากเนื้อความในจดหมายไม่มีข้อความอะไรเป็นสาระมากพอให้สงสัยว่า อีเมลนั้นถูกส่งมาด้วย โปรแกรมไวรัส โดยอาศัยรายชื่ออีเมลที่คนนั้นเก็บไว้ในเครื่อง



อีเมลหลอกล่มมักจะมี link ประหลาดให้กด ให้ตรวจสอบดู address ที่เห็นให้ดีก่อนที่จะกด link ไป เพราะหากกดไปแล้ว บางครั้ง โปรแกรม Antivirus ที่มีอยู่ก็ไม่สามารถกำจัดไวรัสนั้นออกไปได้ บางกรณี หากสังเกตจากเนื้อความก็จะเห็นว่าเป็นจดหมายหลอกล่ม เช่นข้อความข้างล่างนี้ที่อ่านดูจะเห็นว่าเป็นภาษาไทยที่ได้จากการใช้เครื่องแปลหลอกให้ผู้ใช้ให้ข้อมูล user name และ password

เราอยากจะแจ้งให้ทราบว่าขณะนี้เรากำลังดำเนินการบำรุงรักษาที่กำหนดและอัปเดต การให้บริการเว็บของเราและเป็นผลจากการนี้ HTK4S ไวรัสได้รับการตรวจพบใน โฟลเดอร์ที่บัญชีของคุณและบัญชีของคุณจะต้องมีการอัปเดตใหม่ของเรา F-Secure รุ่น HTK4S anti-virus/anti-Spam 2013 เพื่อป้องกันไม่ให้เกิดความเสียหายต่อไฟล์สำคัญของคุณ เติมคอลล์มันด้านล่างและส่งกลับมาหรือบัญชีอีเมลของคุณจะถูกระงับชั่วคราวจากการให้บริการของเรา

ชื่อผู้ใช้:

รหัสผ่าน:

วันเดือนปีเกิด:

***** *

ล้มเหลวที่จะทำเช่นนี้ภายใน 24 ชั่วโมงทันทีจะทำให้บัญชีอีเมลของคุณปิดการใช้งาน
จาก chula.ac.th ในฐานะข้อมูลของเรา,
ลิขสิทธิ์ 2013 chula.ac.th

(ค) สิทธิเครือข่ายทั้งหมด

อย่างไรก็ตาม การติดไวรัสอาจมาจากการใช้เครื่องสาธารณะที่มีผู้อื่นนำไฟล์ที่ติด
ไวรัสมาใช้ เมื่อเรานำเอา flash drive ไปใช้ก็สามารถติดไวรัสมาได้ flash drive จึงไม่ควร
ใช้เพื่อเก็บข้อมูลสำคัญอย่างถาวร ควรคิดเสมอว่าข้อมูลใน flash drive จะถูกลบเมื่อใด
ก็ได้

เมื่อติดไวรัสแล้ว ให้ลองใช้โปรแกรม Antivirus ที่มีอยู่ scan และทำลายไวรัสนั้น
หากไม่ได้ผลให้ลองใช้ windows restore ไปยังวันที่เครื่องไม่มีปัญหา หากยังไม่สามารถ
แก้ไขได้ ให้ใช้โปรแกรม restore image ของ hard drive กลับคืนไป
นอกจากเรื่องไวรัส การถูกจู่โจม โดย hacker ก็เป็นเรื่องที่อาจเกิดขึ้นได้ ทั้งการ hack เพื่อ
ยึดเอา email, facebook, twitter ของเราไปใช้ ทำให้ข้อมูลต่างๆ อาจถูกทำลายได้ สิ่ง
สำคัญคือ ควรมีการเปลี่ยน password อยู่เสมอ อย่าใช้ password ซ้ำเดียวกันกับหลายๆ
account อย่าตั้ง password ที่ง่ายเกินไป สามารถเดาได้ อย่าตั้งไว้สั้นเกินไป มีผู้แนะนำว่า
ให้ใช้ password ยาวแต่จำได้ง่าย อาจใช้คำในภาษามาผสมกันแต่ไม่เป็นวลีที่ใช้ใน
ภาษาเพื่อช่วยให้จำง่าย เช่น “microwave someone Turkish” เป็นต้น

มีตัวอย่างนักคอมพิวเตอร์ ที่ถูก hack Google account, Twitter, Apple ID โดยอาศัย
ช่องโหว่ของระบบการตรวจสอบ โดยได้โทรศัพท์ไปที่ Amazon ขอเพิ่มเลขบัตรเครดิตใน
account โดยมีข้อมูลชื่อ ที่อยู่ และ email เท่านั้น หลังจากนั้น โทรไปหา Amazon ว่าลืม
password และใช้ข้อมูลชื่อ email และหมายเลขบัตรเครดิตที่เพิ่งเพิ่มเข้าไป เมื่อสามารถ
เข้า account Amazon ก็จะสามารถเห็นเลขสี่ตัวสุดท้ายบัตรเครดิตแท้จริง แล้วใช้เลขสี่หลักนี้
โทรไป Apple เพื่ออ้างเป็นเจ้าของ AppleID และขอ password ใหม่ เมื่อยึด AppleID ac-
count ได้ก็สามารถใช้โปรแกรม Find myiPhone เข้าไปลบข้อมูลทั้งหมดใน iPhone, iPad
และ MacNook จากนั้นใช้ข้อมูลนี้ขอ password ของ Gmail ใหม่ และขอ password ของ
Twitter ใหม่ได้ เนื่องจาก setup ให้ขอ reset password โดยส่งไปที่อีก email ที่เรากำหนด

ไว้ได้ (ดู

<http://www.wired.com/gadgetlab/2012/08/apple-amazon-mat-honan-hacking/>)