Cách Import Nutch vào Eclipse

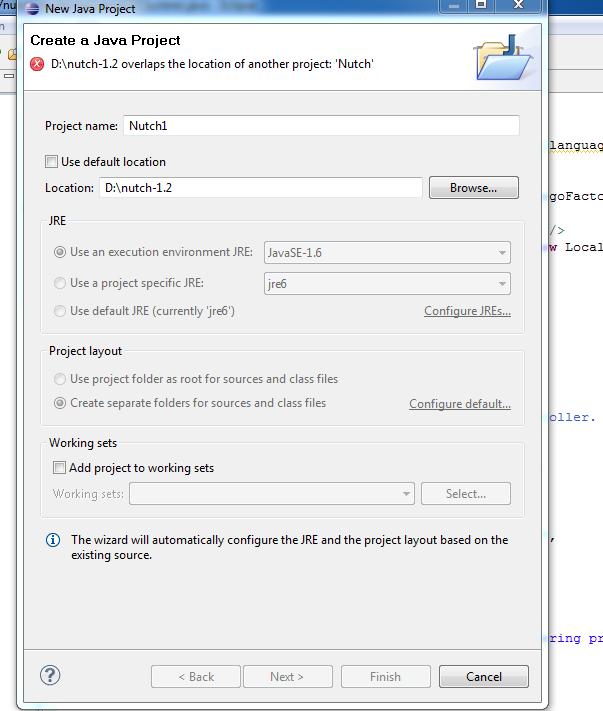
Import nutch vào eclipse thì ta có thể trực tiếp chỉnh sửa core của nutch, sau đó có thể run luôn bằng eclipse.

Nguyên liệu :

* Nutch :D ( phải là bản src nhé ☺ <http://mirror-fpt-telecom.fpt.net/apache//nutch/apache-nutch-1.2-src.zip> )
* Eclipse ( bản nào cũng đc :D, em dùng bản này : [Eclipse IDE for Java Developers](http://www.eclipse.org/downloads/packages/eclipse-ide-java-developers/heliossr1) )
* Cygwin

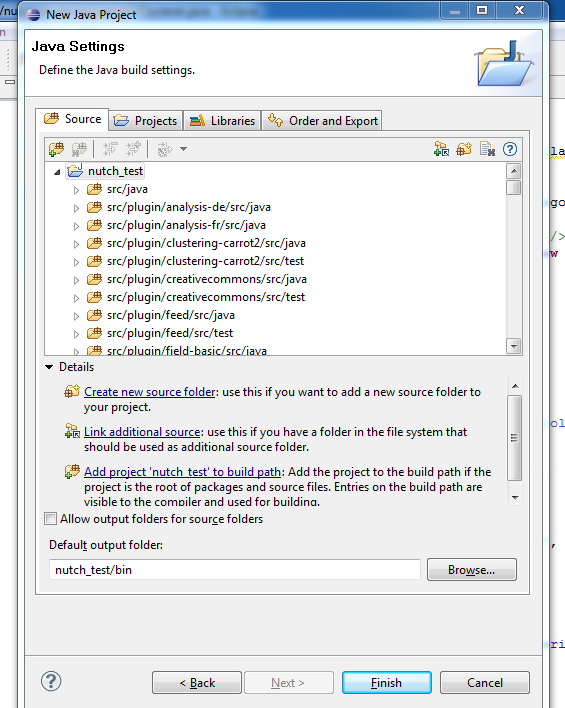
Ở đây là em hướng dẫn import nutch để chạy crawler trực tiếp trên eclipse, ko có hướng dẫn cài web của nutch, ai thích thì cài them vào :D.

Bước 1 : mở eclipse lên, chọn new / java project



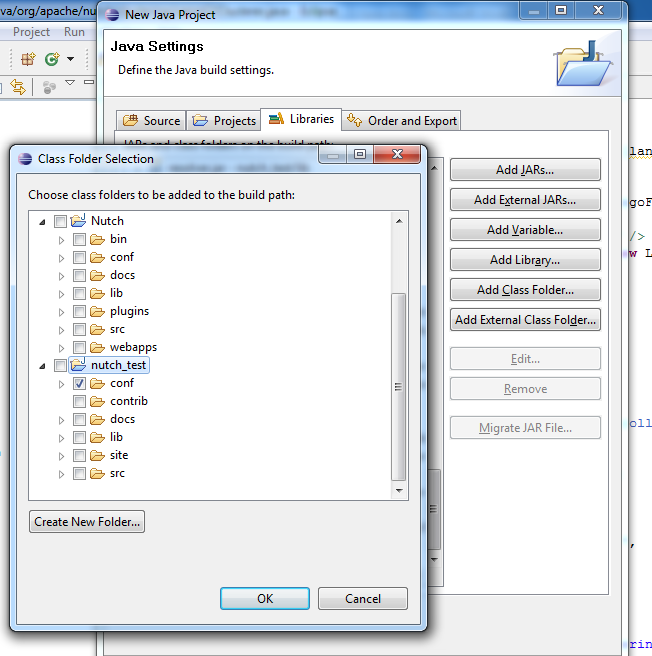
Phần Project name thì để là gì cũng đc, nhưng phần location thì phải là thư mục chứa nutch. Next

Bước 2



Cái default output folder thì để là (tên project)/bin/tmp\_build ( thư mục tmp\_build thì vào nutch tự tạo bằng tay :D )

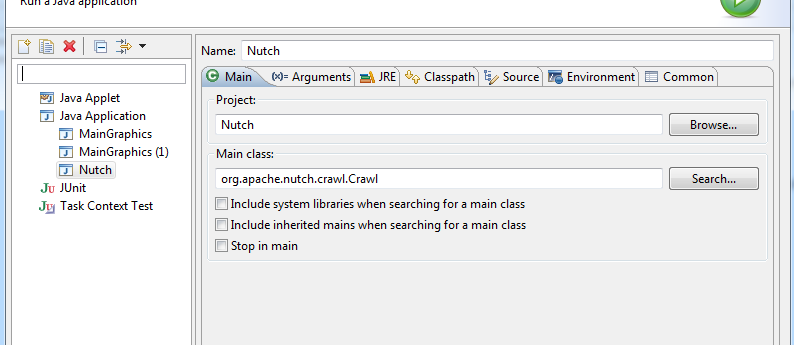
Chọn tab libraries/ add class folder/ tick vào thư mục conf/ ok

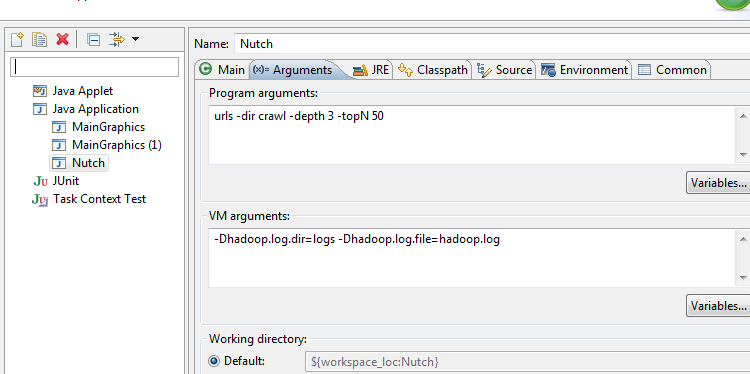


Chọn tab order and export chọn cái thư mục conf vừa add, click top ( để cho nó lên đầu )

Bước 3 : finish :D, nó hỏi gì thì chọn no, ko thì thôi :D

Bước 4 : ở giao diện của eclipse, chọn menu run -> run configurations -> chọn new rồi config như hình





Bước 5 : cài biến môi trường cho cygwin

Vào phần cài đặt biến môi trường, phần name : PATH

Phần content : C:\cygwin\bin

Bước 6 : config cho nutch

Làm theo phần 3 ( phần **Setup the crawler** ) ( chỉ phần 3 thôi nhé :D ) cái 3a ko làm cũng đc :D

<http://zillionics.com/resources/articles/NutchGuideForDummies.htm>

Bước 7 : tắt đi bật lại eclipse và run crawl bình thường :D