Installing Nutch and Solr

Basically what Nutch does is it crawls through a website(link which will be provided by us) , It collects all the data received from the (website/websites/xml's/pdf's) then later it can be saved to the SOLR, which is a search server where you can use to query the information you gave to it through your crawls. So both of these tools work in tandem.

First install **NUTCH** by following **NUTCH_TUTORIAL_1** file given in the references folder. Follow the tutorial step by step and when you install make sure people with less experience only **INSTALL THE BINARY** distribution(instead of the SRC installation) it will make your life much easier.

Secondly install SOLR from **NUTCH_TUTORIAL_1** file itself(solr installation is towards the second half of the file) HERE ALSO MAKE SURE TO DOWNLOAD BINARY FILE itself(will be specified in the file name)

From the step **6** of **NUTCH_TUTORIAL_1** you can integrate nutch and solr together.

SOLR_TUTORIAL_1 contains an elaborate tutorial on how to index files in solr, please have a read after successful installation and a base level testing of both.

Make sure you set your environment variables correctly

If the installation throws any exception take the error part and google it, you can find a lot of help