Nutch and Solr Installation

1-Copy this word file to the virtual machine in home/Lucene open it from there

2- Copy the files in the folder setup04\_files in Lucene folder in your vm

3- Now we are installing Nutch and solr and running them to crawl and index data and perform searches on the crawled data

4- Stemming and conflation code are written (morphological Analysis)

**Step1: copy the files and replace them with existing ones**

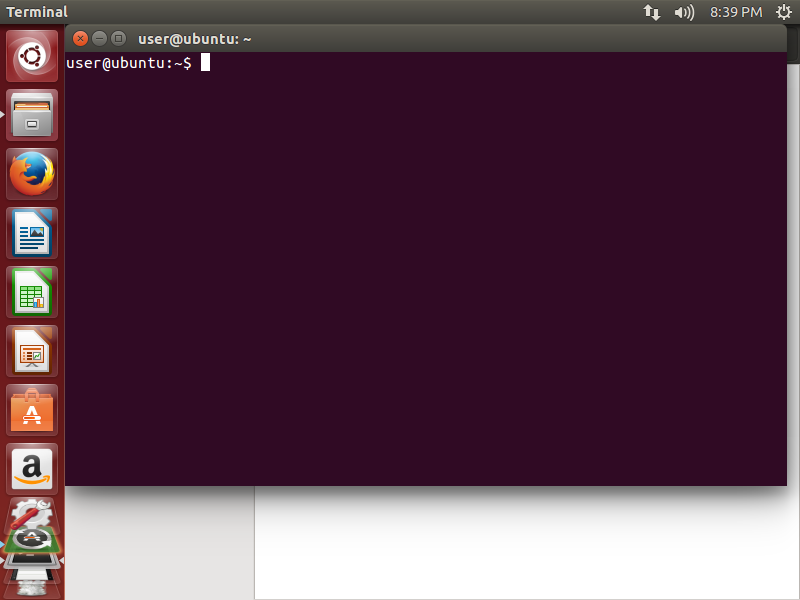
File nutch-site.xml paste it in "Lucene/nutch/conf" (replace existing one)

Folder urls paste it in "Lucene/nutch"

File schema.xml.org and schema.xml and synonyms\_ar.txt and solrconfig.xml paste them in "Lucene/solr/example/solr/collection1/conf/" (replace existing one)

**regex-urlfilter.txt in nutch/conf**

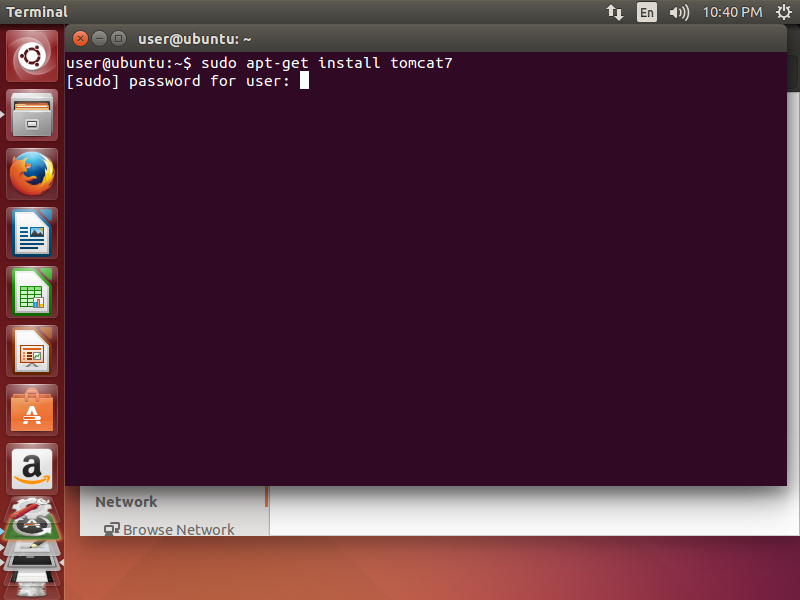
**Step 2:** press ctrl + alt + t opens a new terminal (black screen)



**Step 3:** paste the following lines in the terminal

sudo apt-get install tomcat7

if asks for password : 123

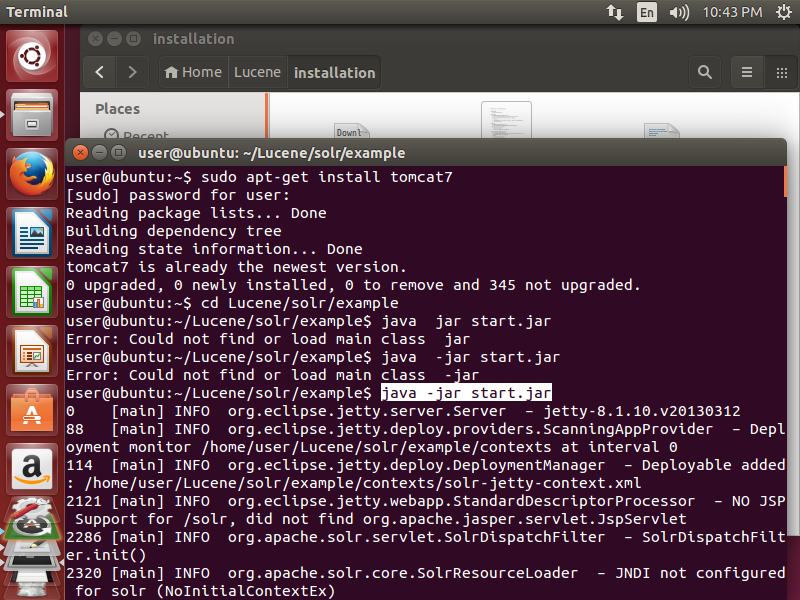


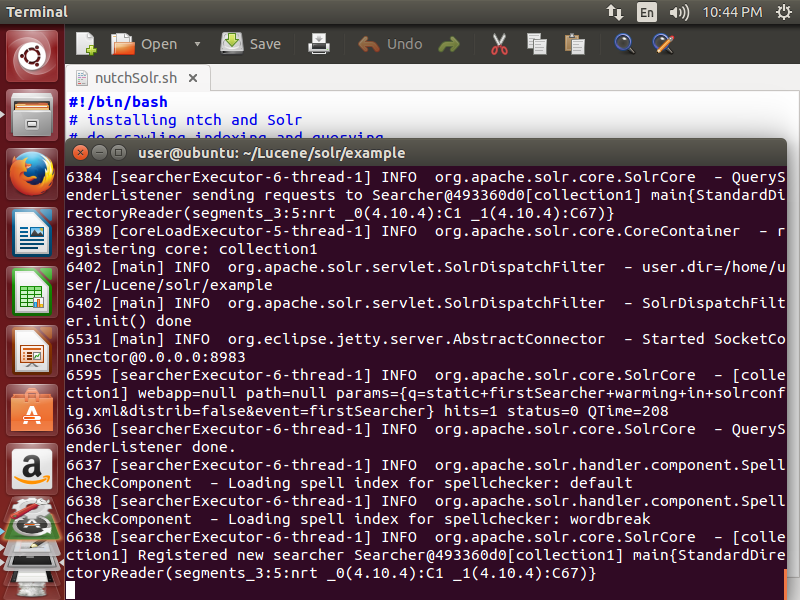
**Step 4:** paste the following

cd Lucene/solr/example

write this with ur hand do not copy and paste

java ­jar start.jar





**Step 5 :** paste the following in new terminal

ctrl + alt + t open new terminal

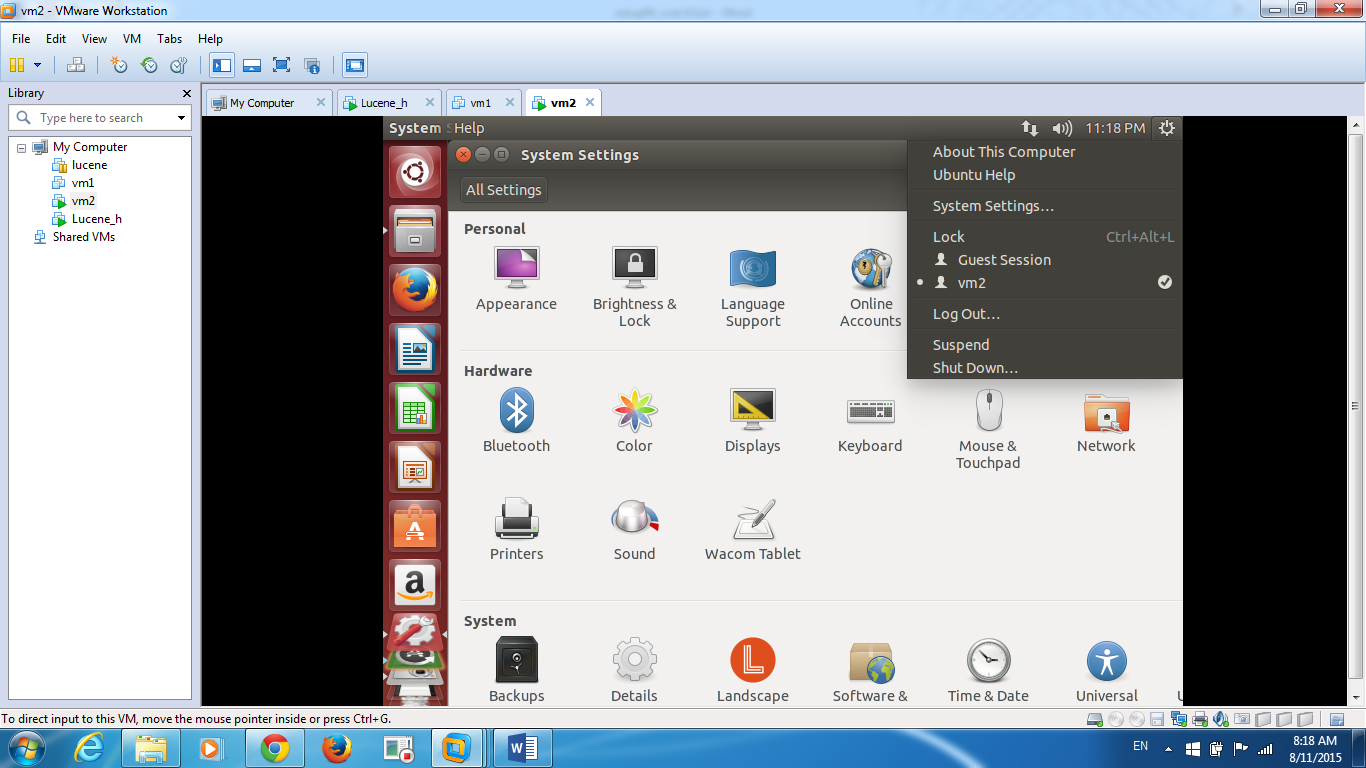
**cd Lucene/nutch**

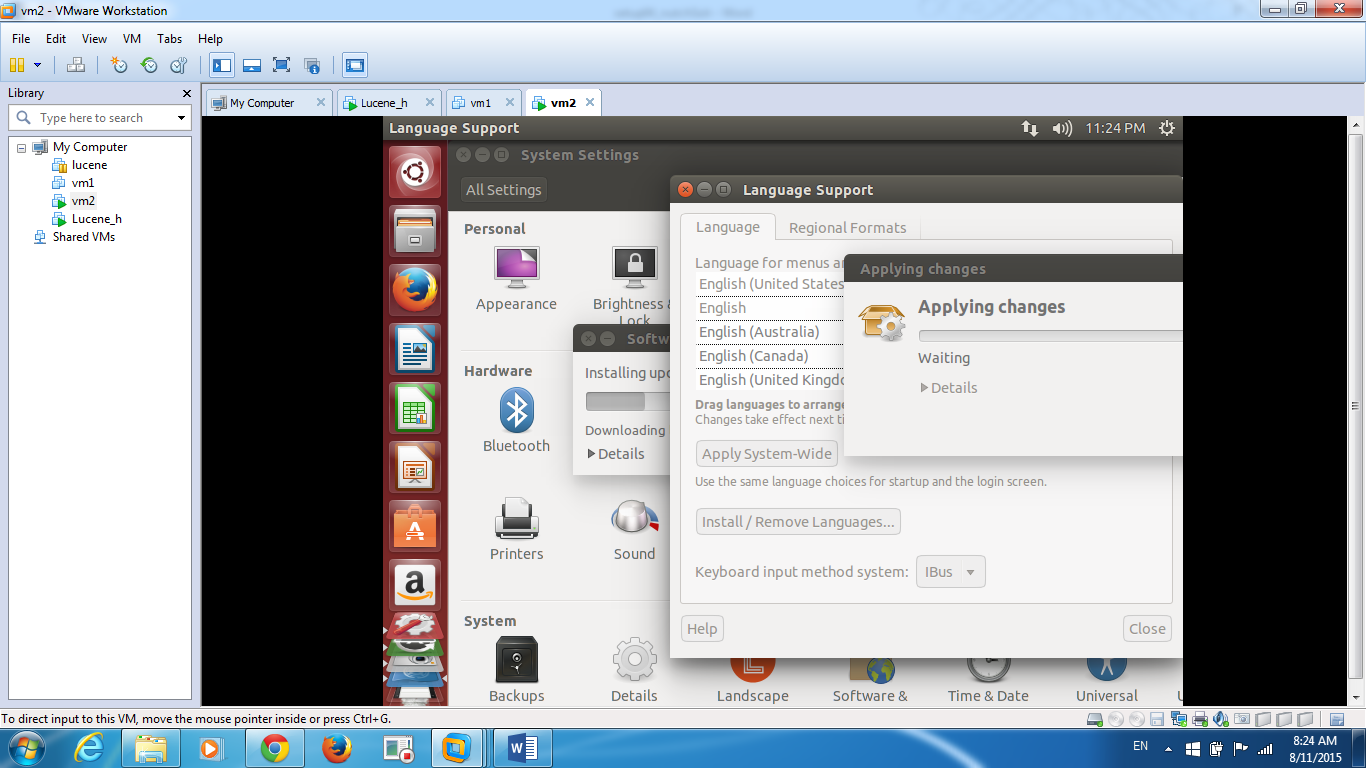
**bin/crawl urls/ crawl/ http://localhost:8983/solr/ 3**

**Crawling**

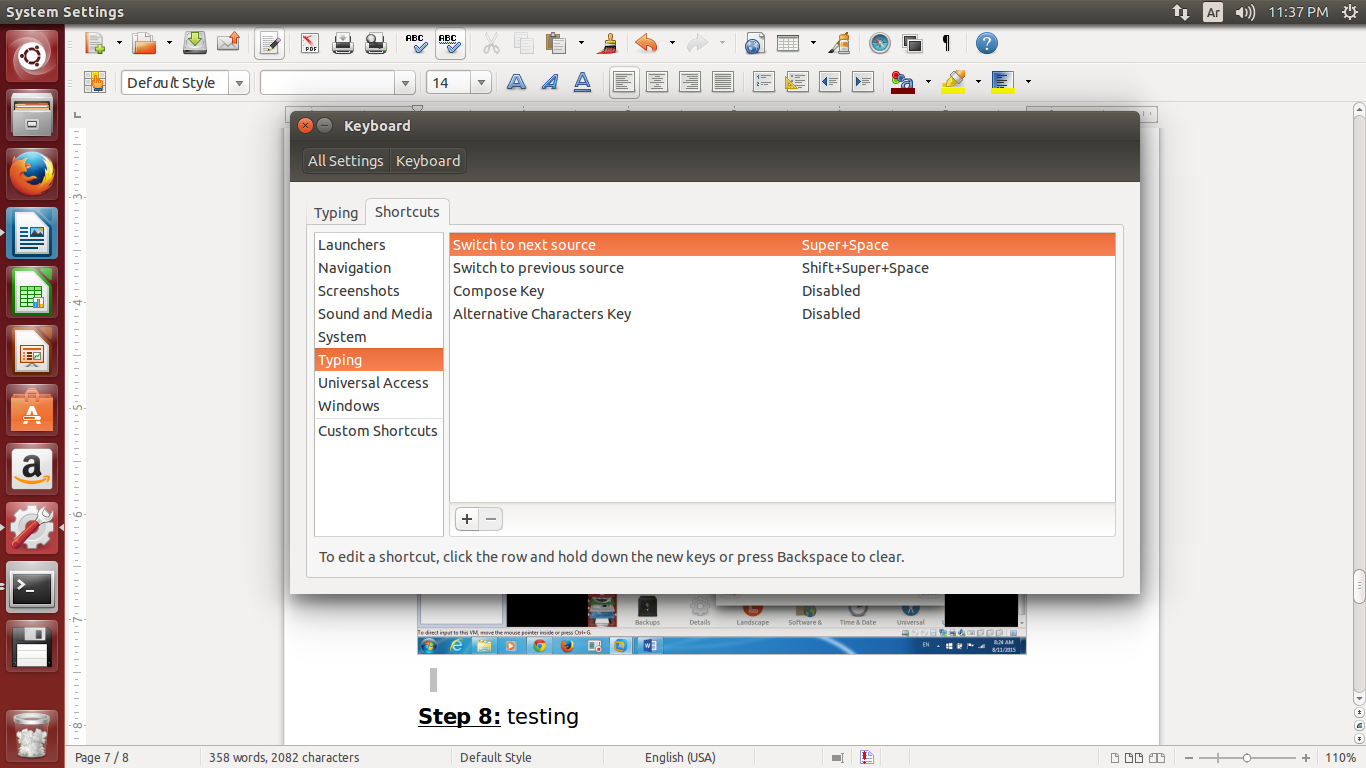
Do not close the terminal

**Step 6:** install Arabic language on linux and configure setting to change langauges

choose system settings 🡪 language support (if asks for password 123)🡪 install and remove language button🡪 scroll and choose Arabic



choose system settings🡪 text entry🡪 press + to add Arabic then press + again to add English 🡪 choose keyboards setting from the right bottom corner



On the first field on the right (Switch to next source) click on it , it will change to new accelerator , press shift + 1

On the second field on the right (Switch to previous source) click on it , it will change to new accelerator , press shift + 2

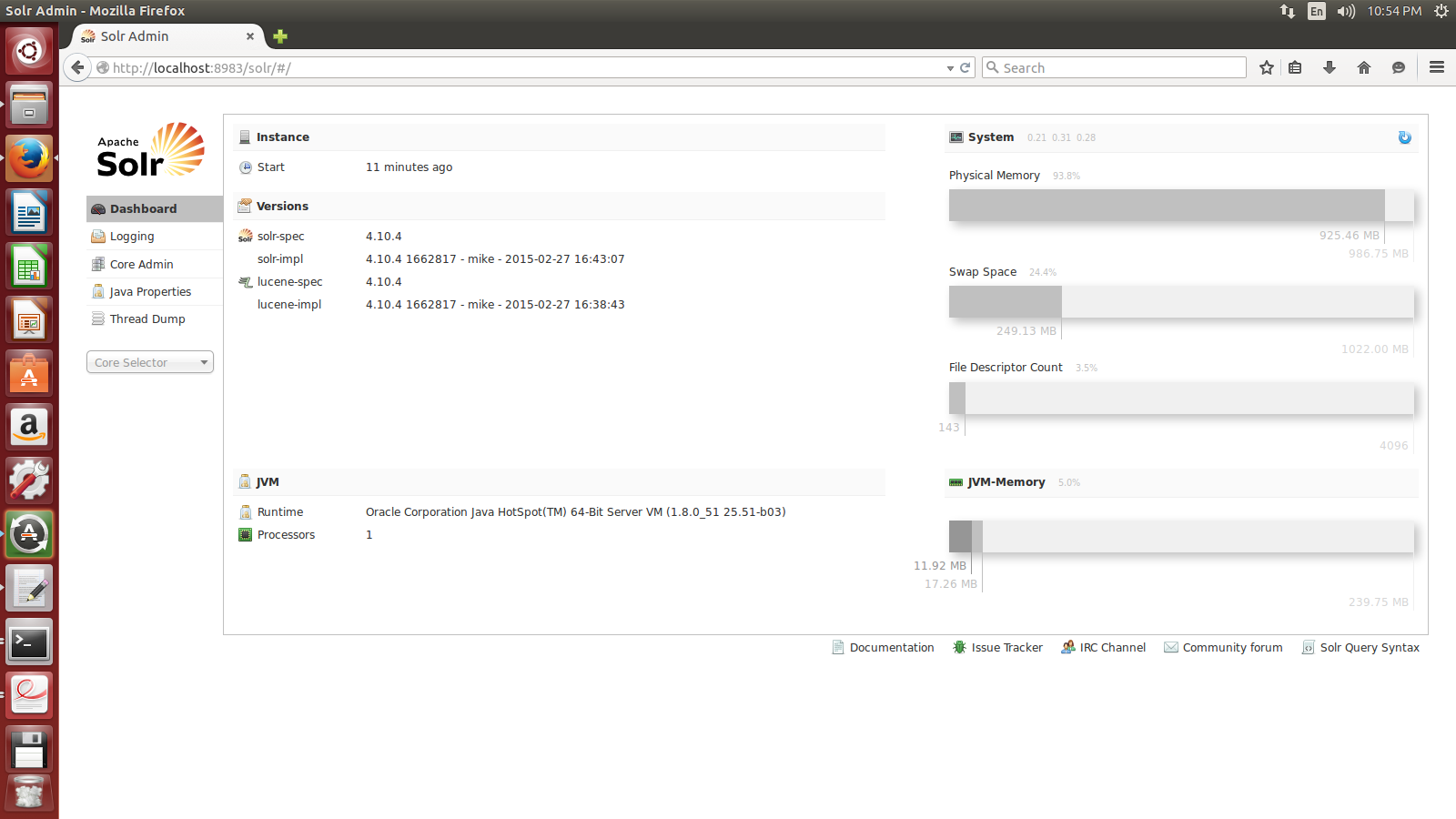
لتغيير بين اللغتينshift +1 او shift +2

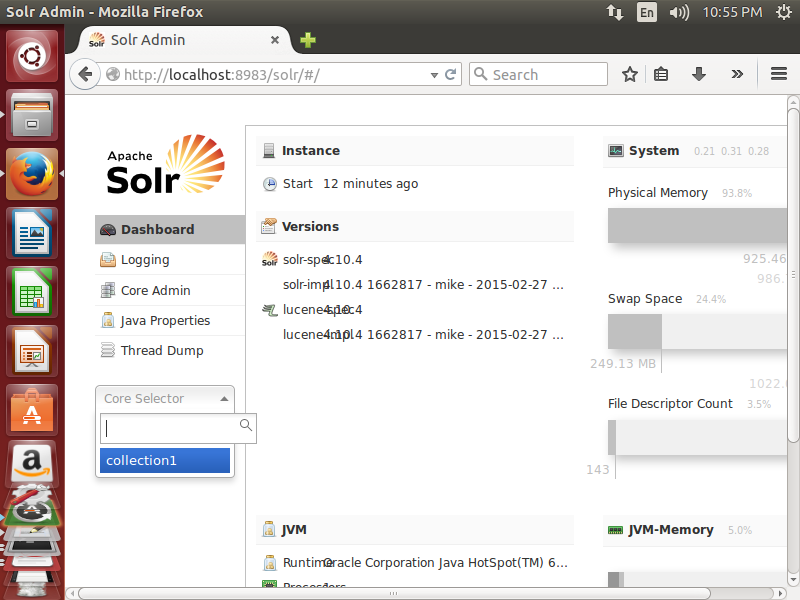
Close everything

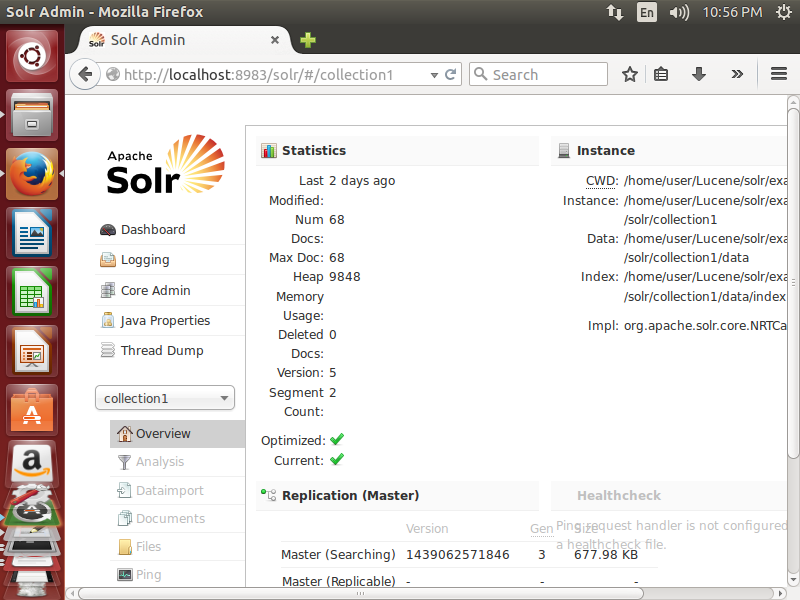
**Step 6: ده خطوة التشغيل querying**

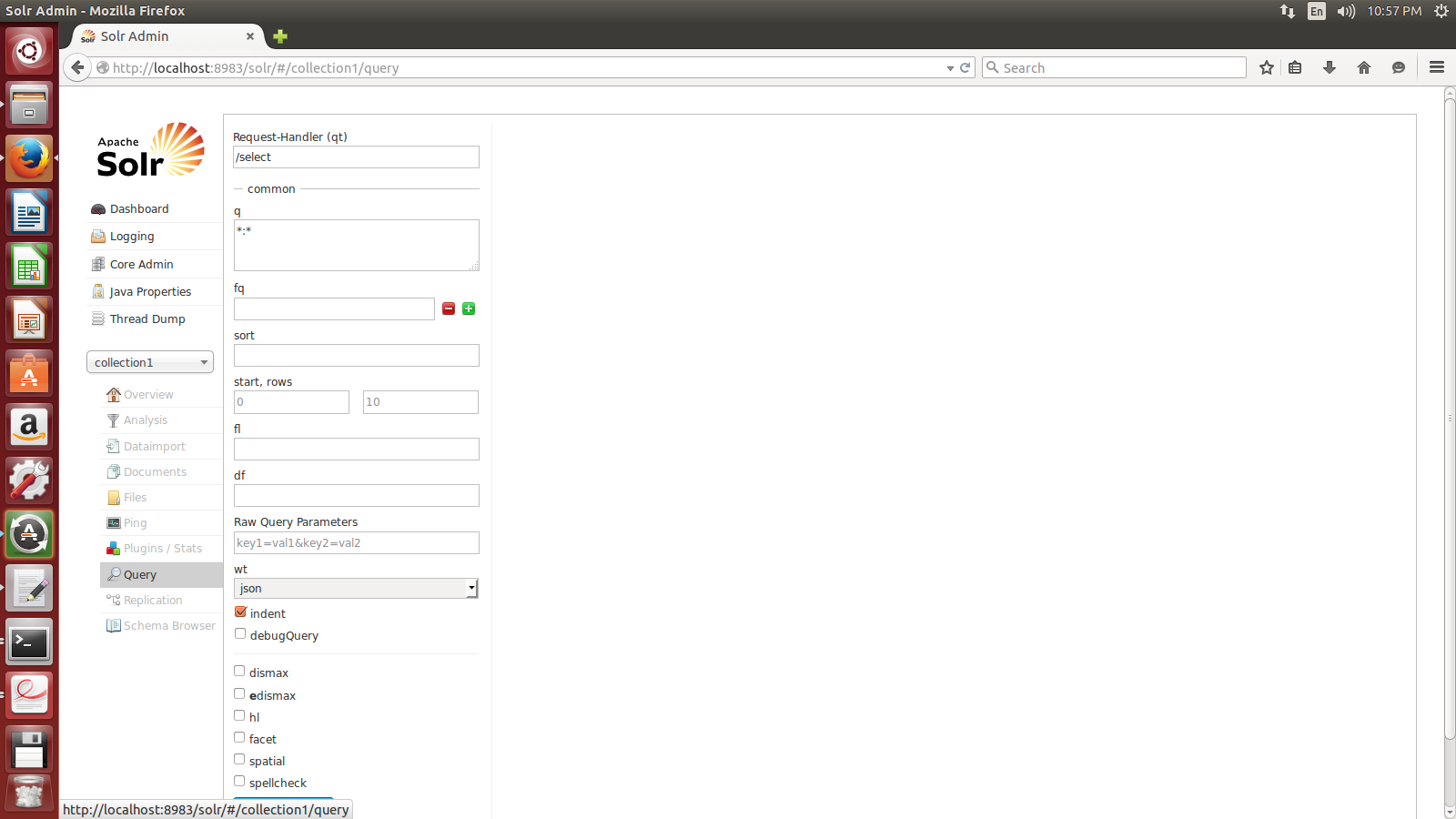
<http://localhost:8983/solr/admin>

This window will appear









Try a query

**Step 8:** testing

Try words testing stemming and synonyms