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

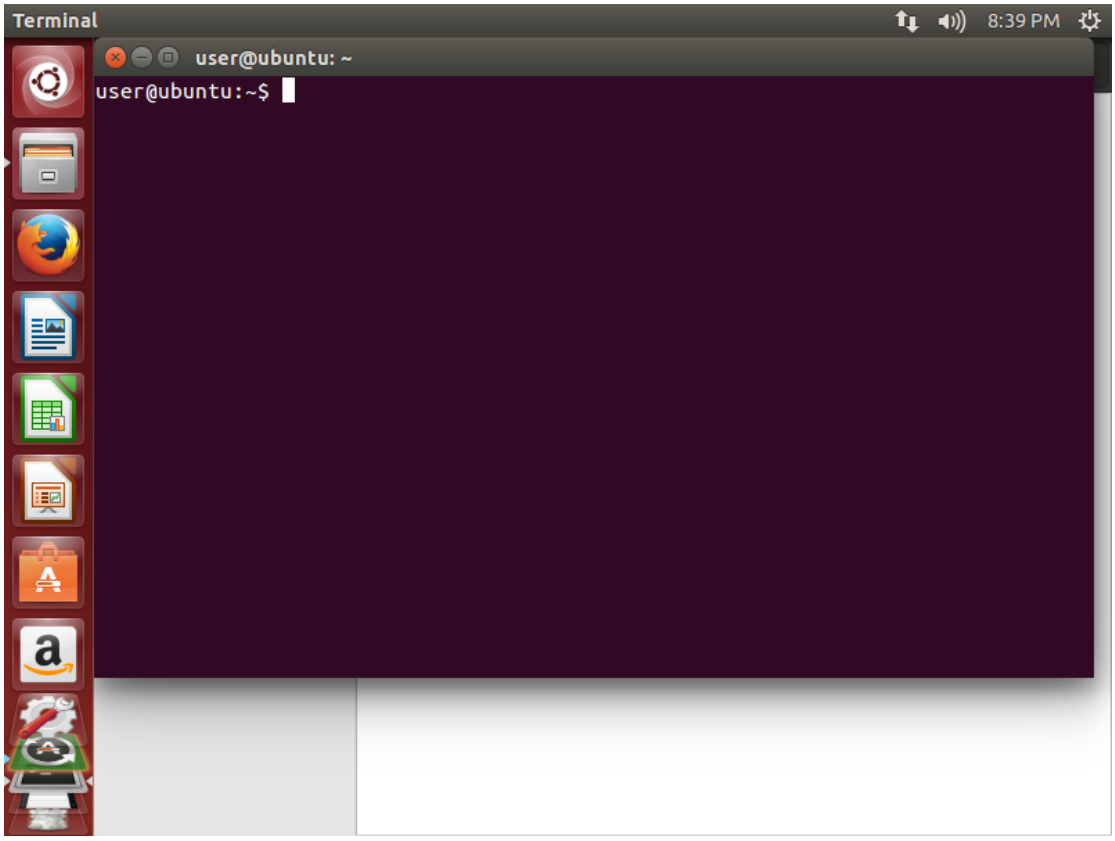
File nutch-site.xml paste it in "Lucene/nutch/conf" (replace existing one)

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

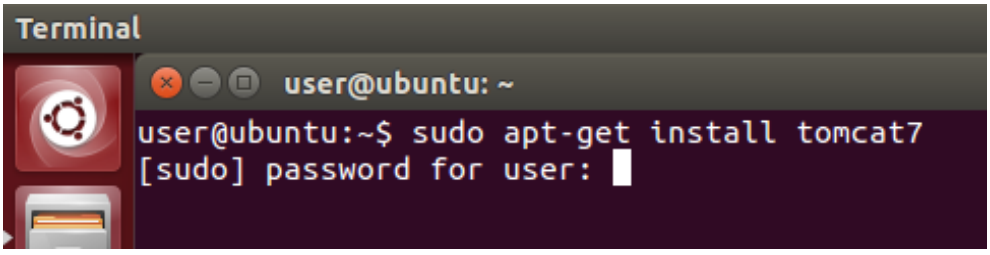
Step 2: press ctrl + alt + t opens a new terminal (black screen)



Step 3: paste the following lines in the terminal

sudo apt-get install tomcat7

if asks for password : 123

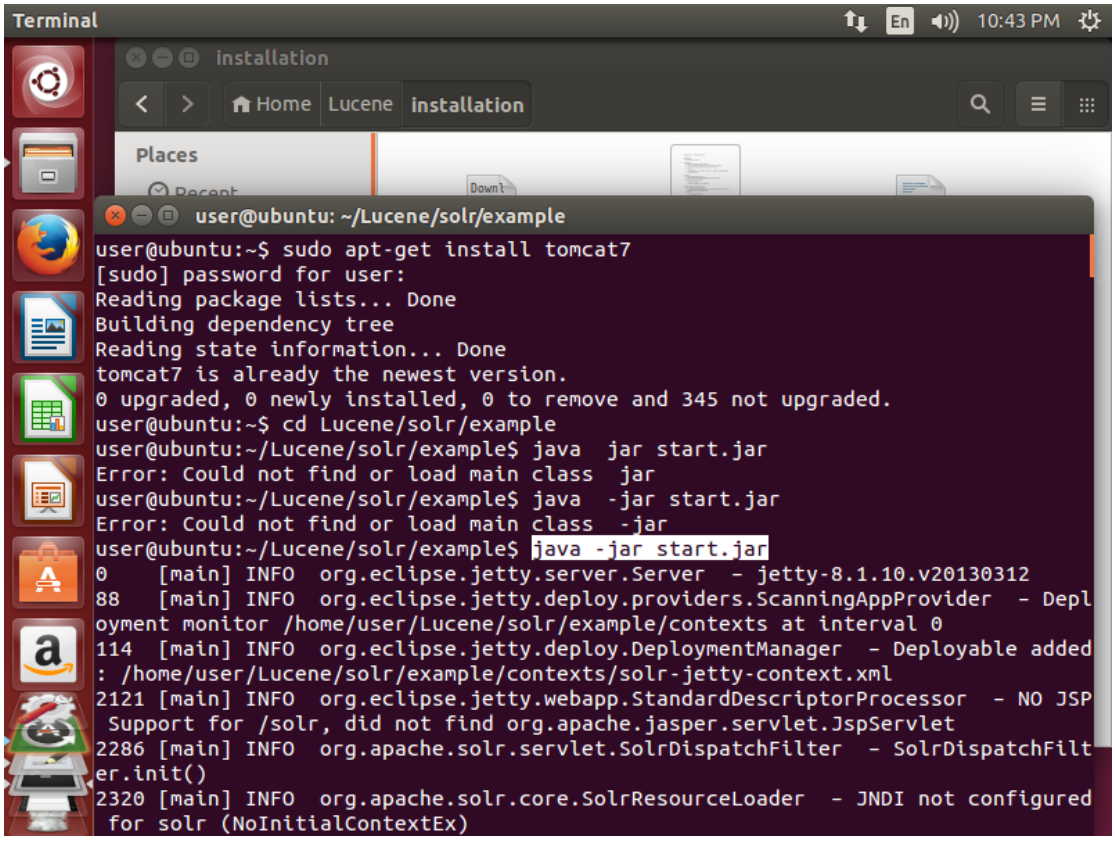


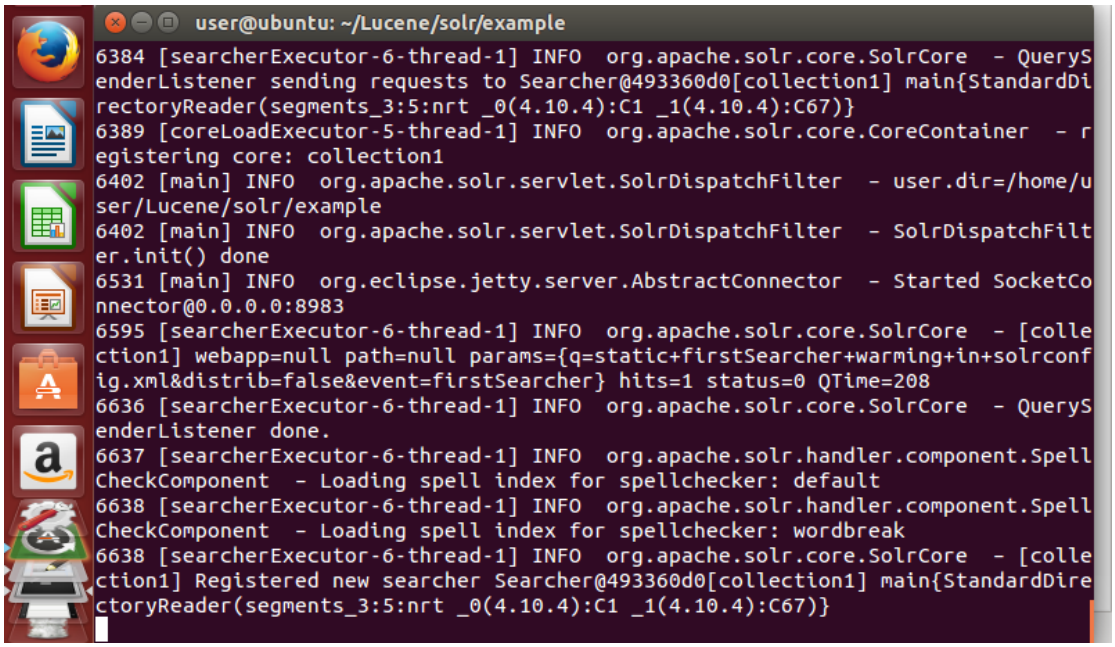
Step 4: paste the following

cd Lucene/solr/example

write this with ur hand do not copy and paste

java -jar start.jar





<u>Step 5 :</u> paste the following in new terminal

ctrl + alt + t open new terminal

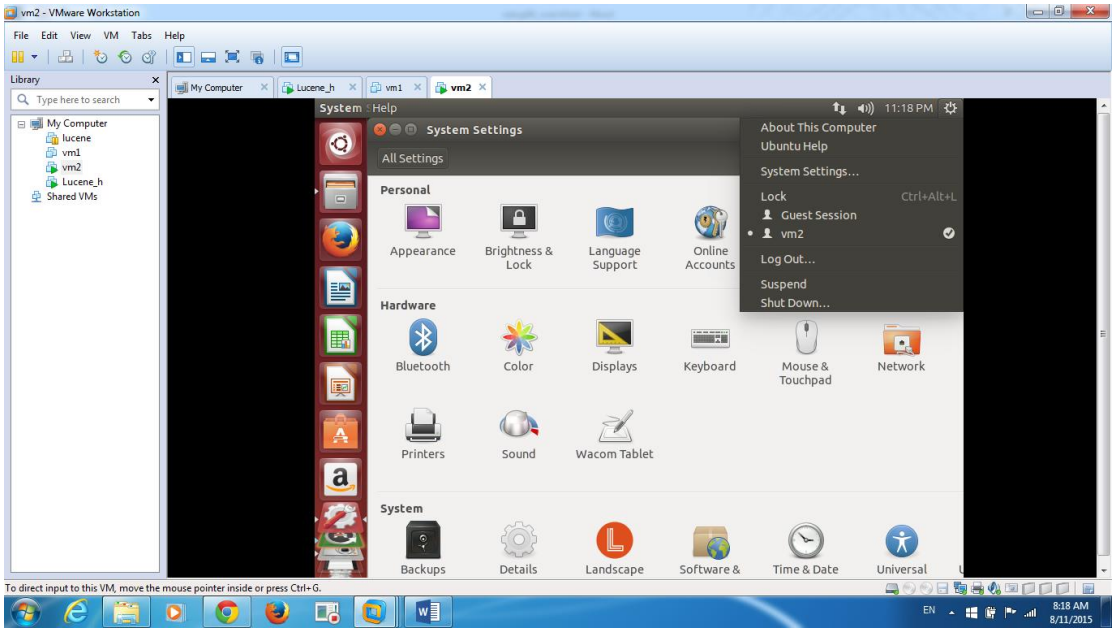
cd Lucene/nutch

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

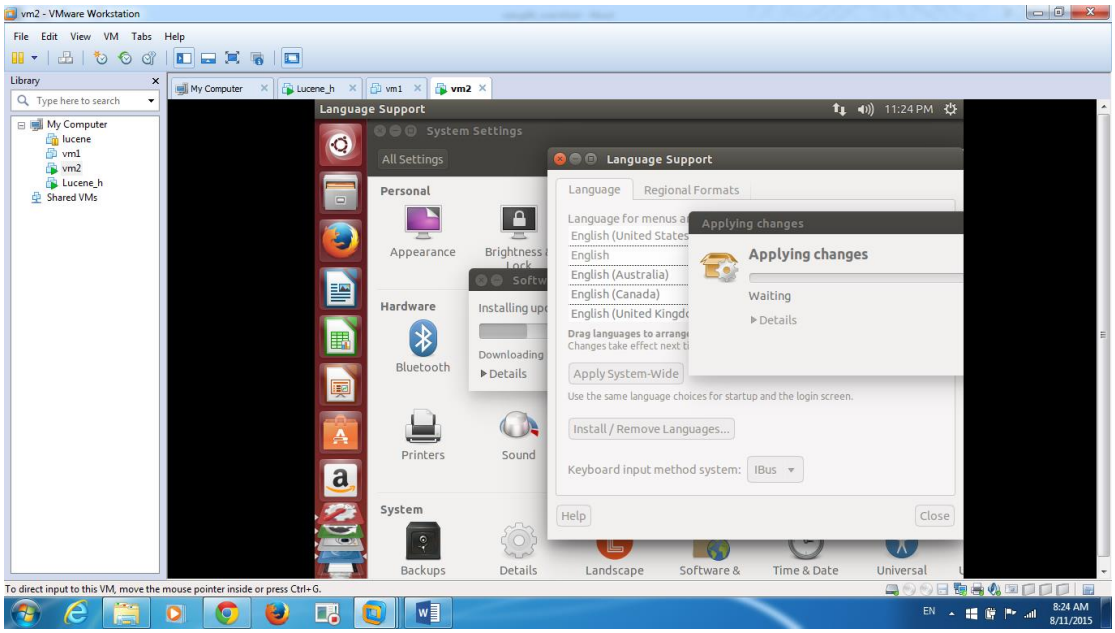
Crawling

Do not close the terminal

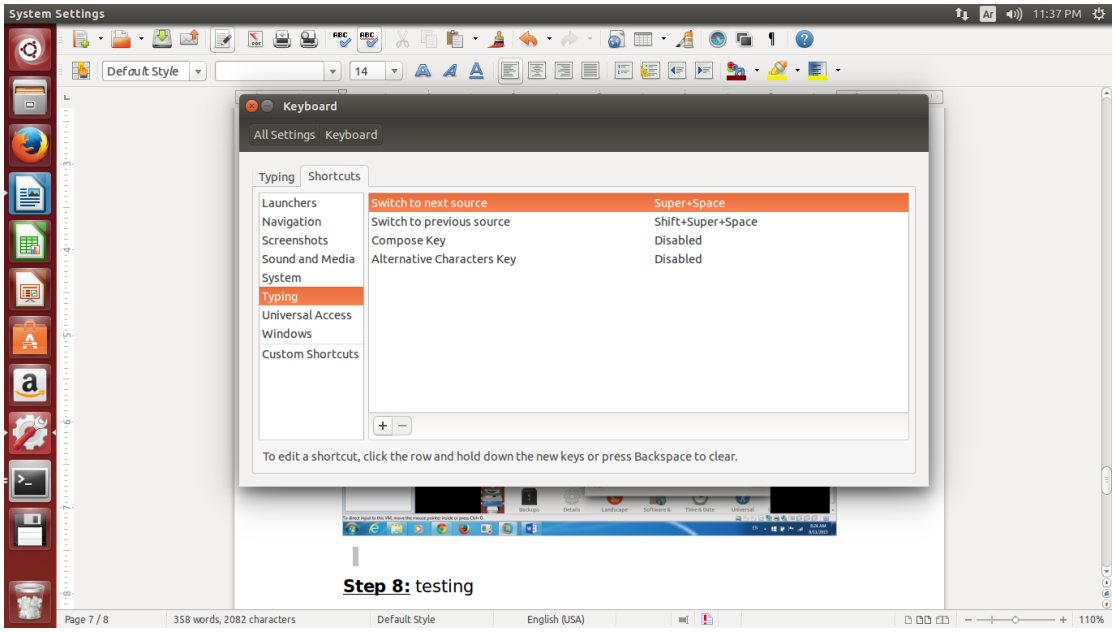
Step 6: install Arabic language on linux and configure setting to change langauges



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

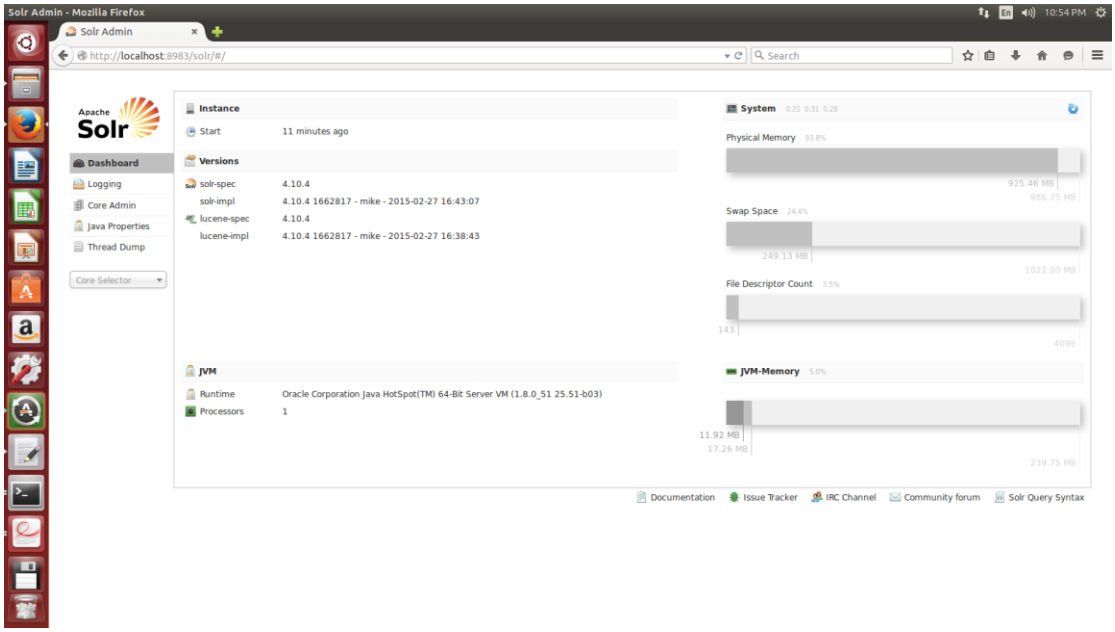
لتغيير بين اللغتين shift +1 او shift +2

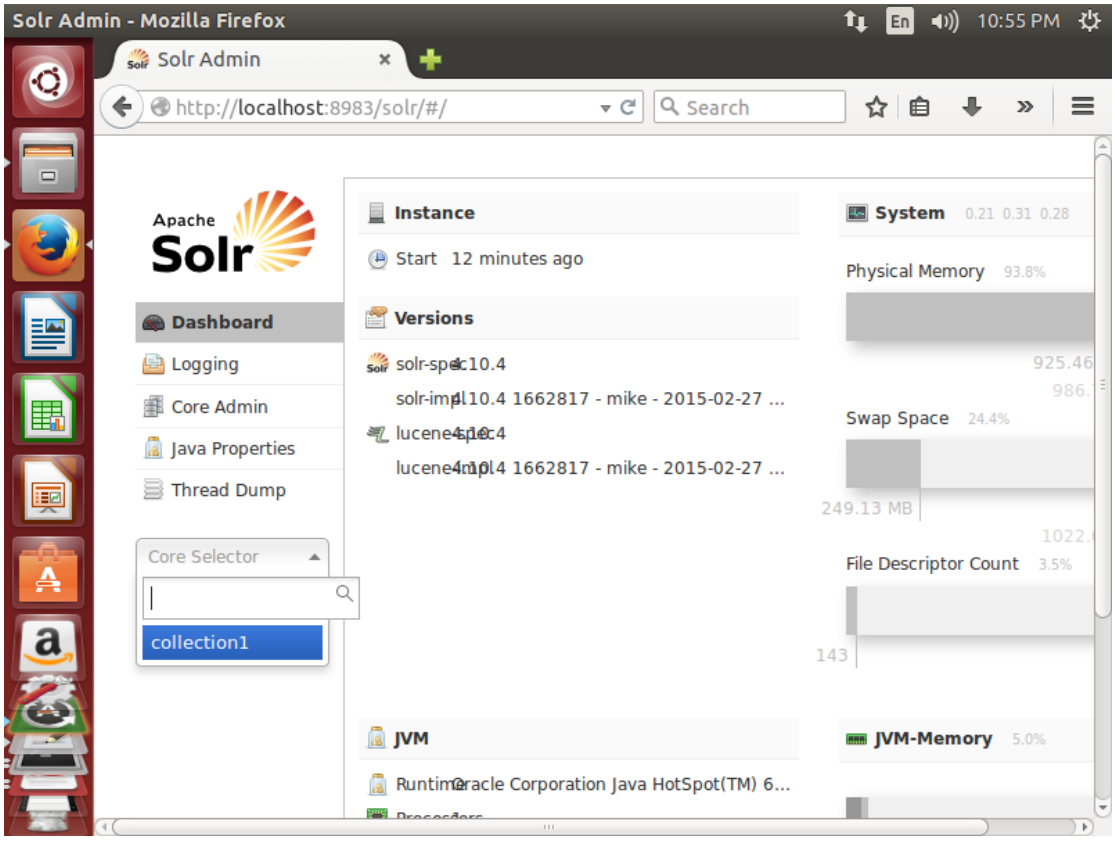
Close everything

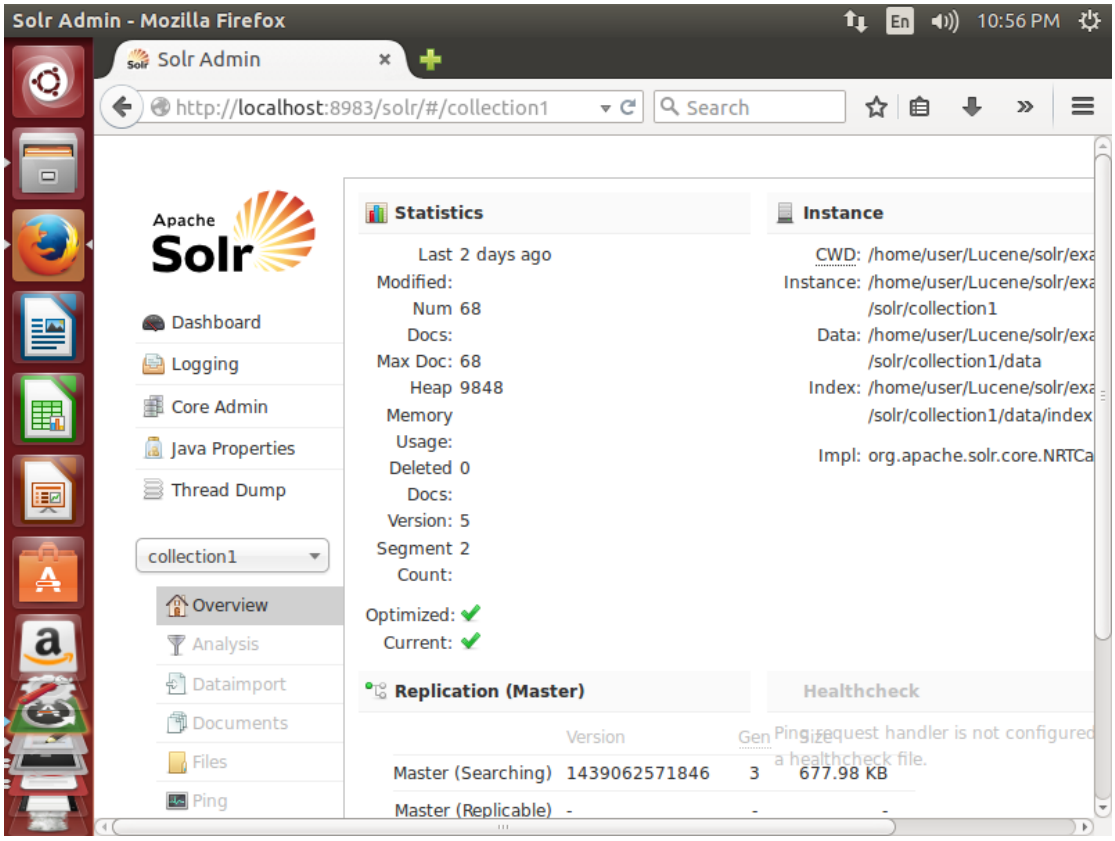
Step 6: ده خطوة التشغيل querying

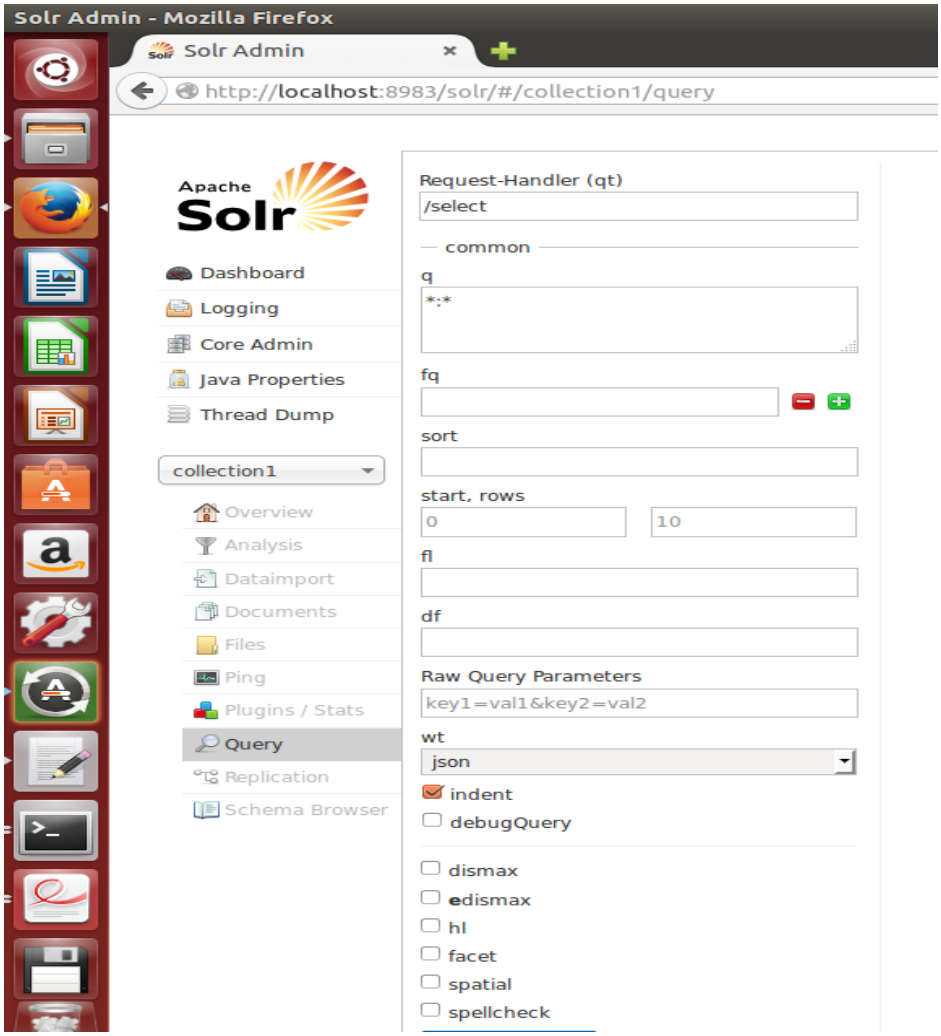
http://localhost:8983/solr/admin

This window will appear


Solr Admin - Mozilla Firefox
Solr Admin
http://localhost:8983/solr/#/
Search
Apache Solr
Dashboard
Logging
Core Admin
Java Properties
Thread Dump
Core Selector
Instance
Start 11 minutes ago
Versions
solr-spec 4.10.4
solr-impl 4.10.4 1662817 - mike - 2015-02-27 16:43:07
lucene-spec 4.10.4
lucene-impl 4.10.4 1662817 - mike - 2015-02-27 16:38:43
System 0.21 0.31 0.28
Physical Memory 93.8%
925.46 MB
986.75 MB
Swap Space 24.4%
249.13 MB
1022.00 MB
File Descriptor Count 3.5%
143
4096
JVM
Runtime Oracle Corporation Java HotSpot(TM) 64-Bit Server VM (1.8.0_51 25.51-b03)
Processors 1
JVM-Memory 5.0%
11.92 MB
17.26 MB
239.75 MB
Documentation
Issue Tracker
IRC Channel
Community forum
Solr Query Syntax


Solr Admin - Mozilla Firefox
10:55 PM
Solr Admin
http://localhost:8983/solr/#/
Search
Apache Solr
Dashboard
Logging
Core Admin
Java Properties
Thread Dump
Core Selector
collection1
Instance
Start 12 minutes ago
Versions
solr-spec 4.10.4
solr-impl 4.10.4 1662817 - mike - 2015-02-27 ...
lucene-spec 4.10.4
lucene-impl 4.10.4 1662817 - mike - 2015-02-27 ...
System 0.21 0.31 0.28
Physical Memory 93.8%
925.46
Swap Space 24.4%
249.13 MB
File Descriptor Count 3.5%
143
JVM
Runtime Oracle Corporation Java HotSpot(TM) 6...
JVM-Memory 5.0%


Solr Admin - Mozilla Firefox
10:56 PM
Solr Admin
http://localhost:8983/solr/#/collection1
Search
Apache Solr
Dashboard
Logging
Core Admin
Java Properties
Thread Dump
collection1
Overview
Analysis
Dataimport
Documents
Files
Ping
Statistics
Last Modified: 2 days ago
Num Docs: 68
Max Doc: 68
Heap Memory Usage: 9848
Deleted Docs: 0
Version: 5
Segment Count: 2
Optimized:
Current:
Instance
CWD: /home/user/Lucene/solr/exa
Instance: /home/user/Lucene/solr/exa /solr/collection1
Data: /home/user/Lucene/solr/exa /solr/collection1/data
Index: /home/user/Lucene/solr/exa /solr/collection1/data/index
Impl: org.apache.solr.core.NRTCa
Replication (Master)
Version
Gen
Size
Master (Searching) 1439062571846 3 677.98 KB
Master (Replicable) - - -
Healthcheck
Ping request handler is not configured a healthcheck file.



Try a query

Step 8: testing

Try words testing stemming and synonyms