

## 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)

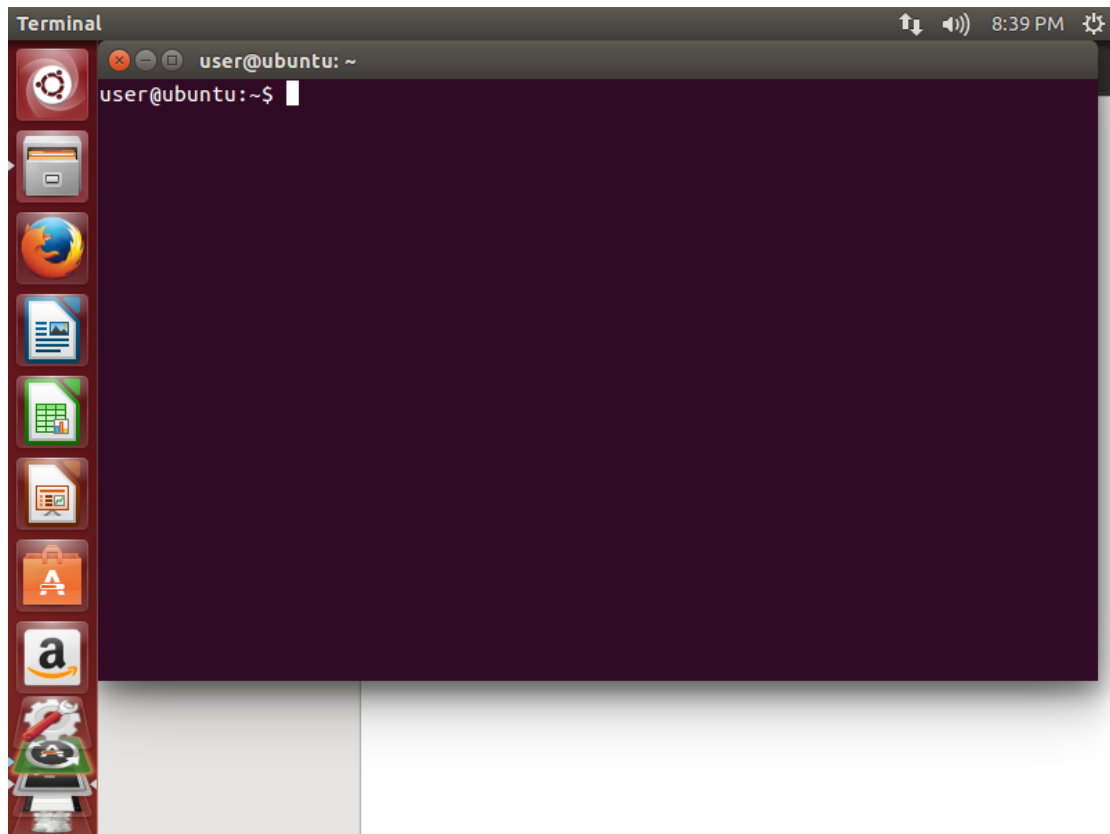
Folder **urls** paste it in "Lucene/nutch"

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**

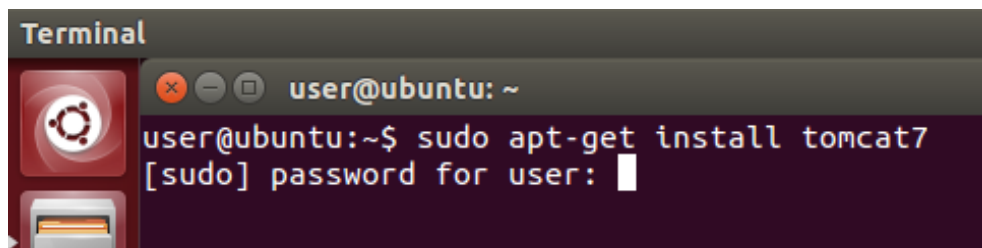
**Step 2:** press ctrl + alt + t opens a new terminal (black screen)



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

`sudo apt-get install tomcat7`

if asks for password : 123

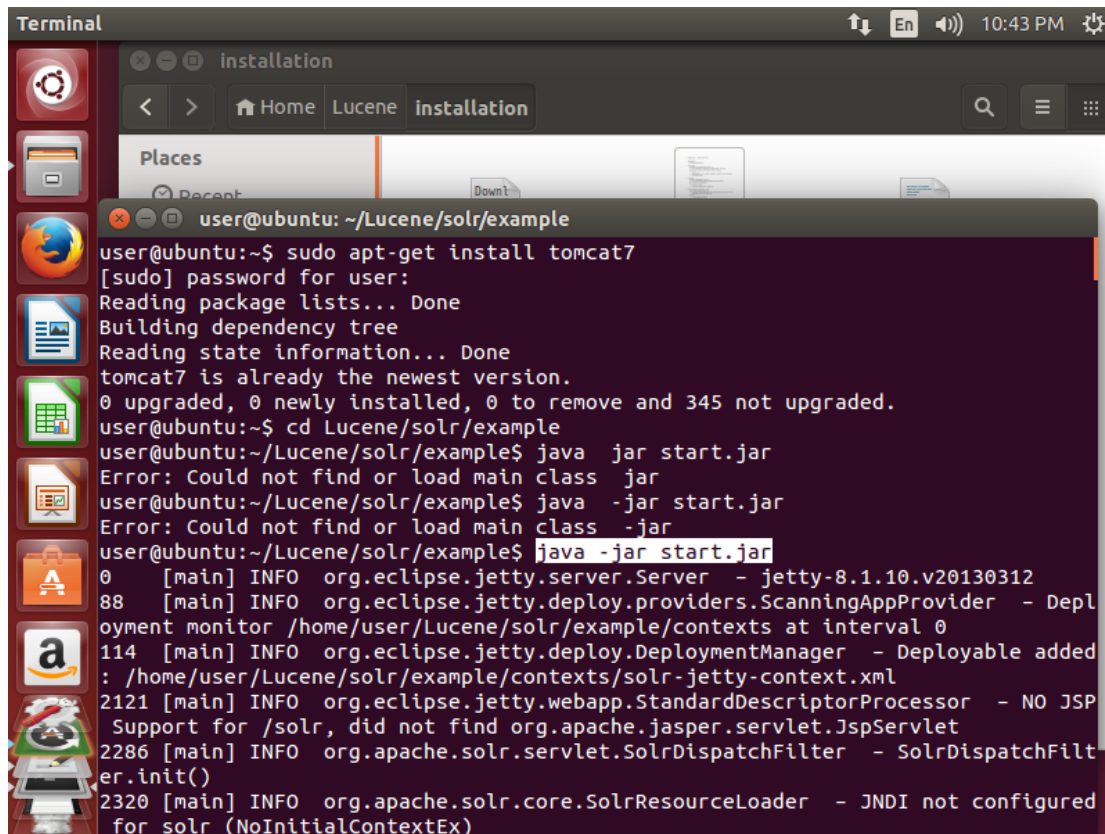


**Step 4:** paste the following

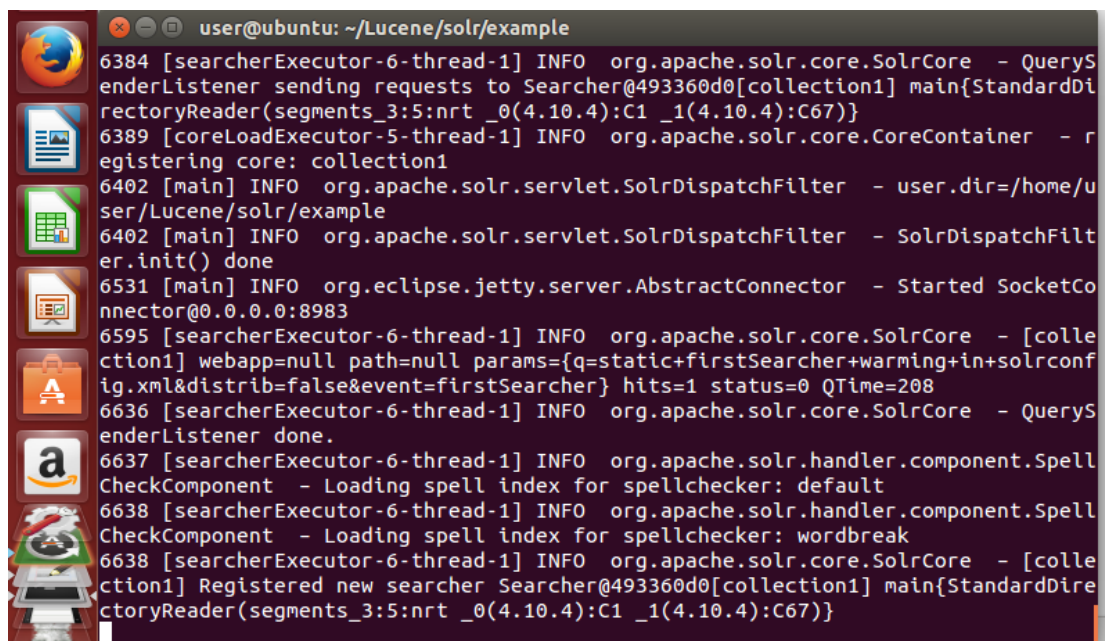
`cd Lucene/solr/example`

write this with ur hand do not copy and paste

`java -jar start.jar`



```
Terminal
installation
user@ubuntu: ~/Lucene/solr/example
user@ubuntu:~$ sudo apt-get install tomcat7
[sudo] password for user:
Reading package lists... Done
Building dependency tree
Reading state information... Done
tomcat7 is already the newest version.
0 upgraded, 0 newly installed, 0 to remove and 345 not upgraded.
user@ubuntu:~$ cd Lucene/solr/example
user@ubuntu:~/Lucene/solr/example$ java jar start.jar
Error: Could not find or load main class jar
user@ubuntu:~/Lucene/solr/example$ java -jar start.jar
Error: Could not find or load main class -jar
user@ubuntu:~/Lucene/solr/example$ java -jar start.jar
0 [main] INFO org.eclipse.jetty.server.Server - jetty-8.1.10.v20130312
88 [main] INFO org.eclipse.jetty.deploy.providers.ScanningAppProvider - Deployment monitor /home/user/Lucene/solr/example/contexts at interval 0
114 [main] INFO org.eclipse.jetty.deploy.DeploymentManager - Deployable added : /home/user/Lucene/solr/example/contexts/solr-jetty-context.xml
2121 [main] INFO org.eclipse.jetty.webapp.StandardDescriptorProcessor - NO JSP Support for /solr, did not find org.apache.jasper.servlet.JspServlet
2286 [main] INFO org.apache.solr.servlet.SolrDispatchFilter - SolrDispatchFilter.init()
2320 [main] INFO org.apache.solr.core.SolrResourceLoader - JNDI not configured for solr (NoInitialContextEx)
```



```
user@ubuntu: ~/Lucene/solr/example
6384 [searcherExecutor-6-thread-1] INFO org.apache.solr.core.SolrCore - QuerySenderListener sending requests to Searcher@493360d0[collection1] main{StandardDirectoryReader(segments_3:5:nrt_0(4.10.4):C1_1(4.10.4):C67)}
6389 [coreLoadExecutor-5-thread-1] INFO org.apache.solr.core.CoreContainer - registering core: collection1
6402 [main] INFO org.apache.solr.servlet.SolrDispatchFilter - user.dir=/home/user/Lucene/solr/example
6402 [main] INFO org.apache.solr.servlet.SolrDispatchFilter - SolrDispatchFilter.init() done
6531 [main] INFO org.eclipse.jetty.server.AbstractConnector - Started SocketConnector@0.0.0.0:8983
6595 [searcherExecutor-6-thread-1] INFO org.apache.solr.core.SolrCore - [collection1] webapp=null path=null params={q=static+firstSearcher+warming+in+solrconfig.xml&istrib=false&event=firstSearcher} hits=1 status=0 QTime=208
6636 [searcherExecutor-6-thread-1] INFO org.apache.solr.core.SolrCore - QuerySenderListener done.
6637 [searcherExecutor-6-thread-1] INFO org.apache.solr.handler.component.SpellCheckComponent - Loading spell index for spellchecker: default
6638 [searcherExecutor-6-thread-1] INFO org.apache.solr.handler.component.SpellCheckComponent - Loading spell index for spellchecker: wordbreak
6638 [searcherExecutor-6-thread-1] INFO org.apache.solr.core.SolrCore - [collection1] Registered new searcher Searcher@493360d0[collection1] main{StandardDirectoryReader(segments_3:5:nrt_0(4.10.4):C1_1(4.10.4):C67)}
```

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

ctrl + alt + t open new terminal

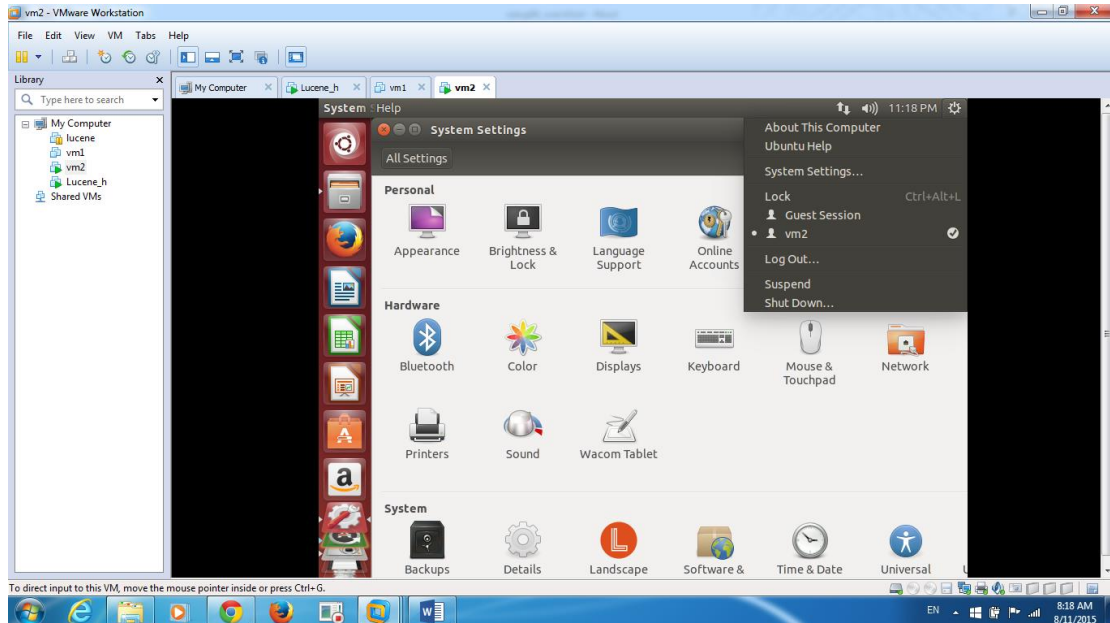
**cd Lucene/nutch**

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

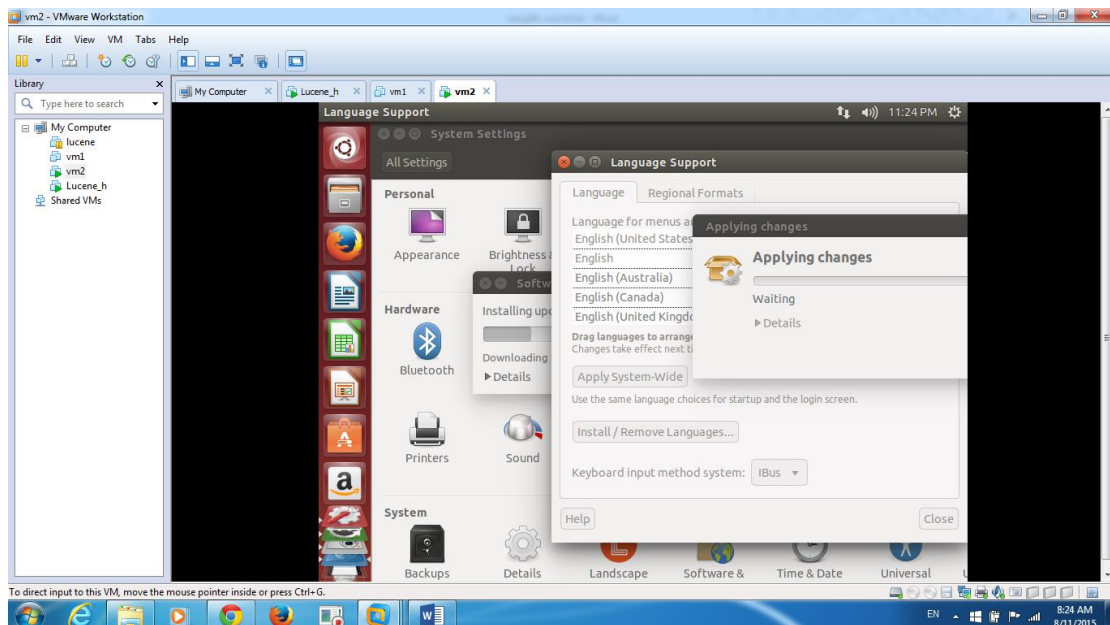
## Crawling

Do not close the terminal

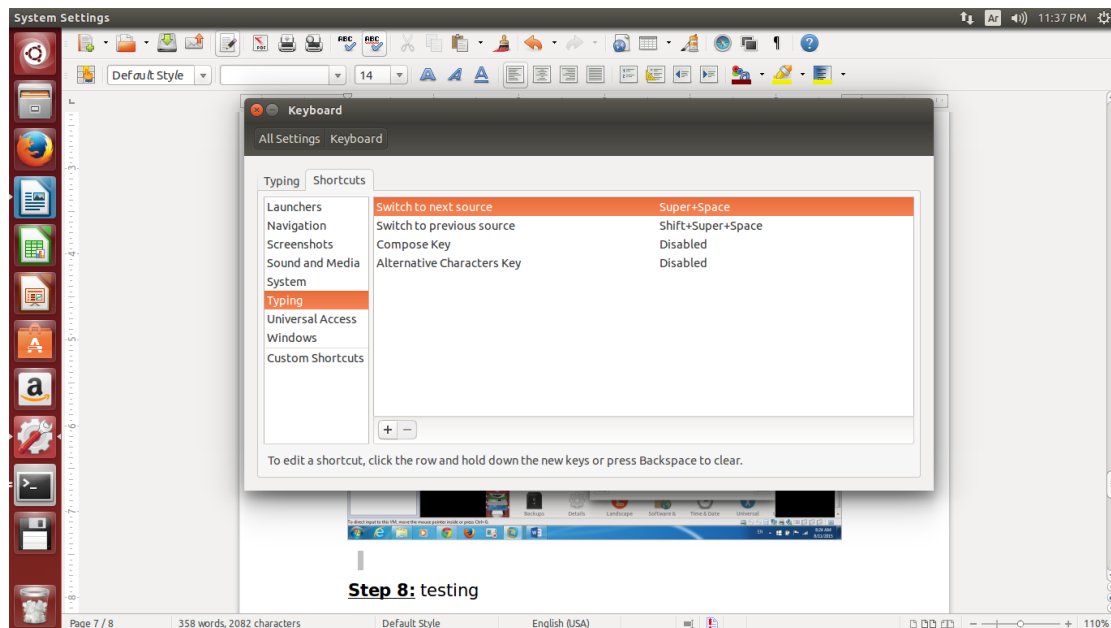
**Step 6:** install Arabic language on linux and configure setting to change languages



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

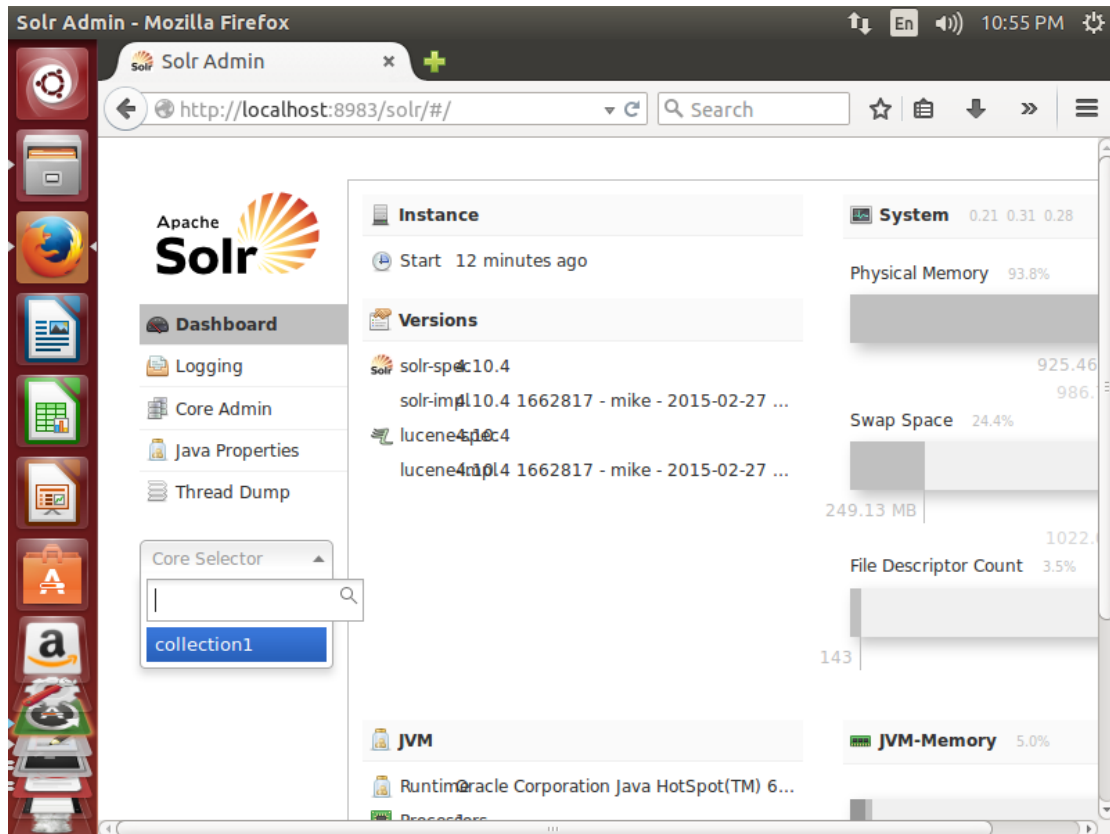
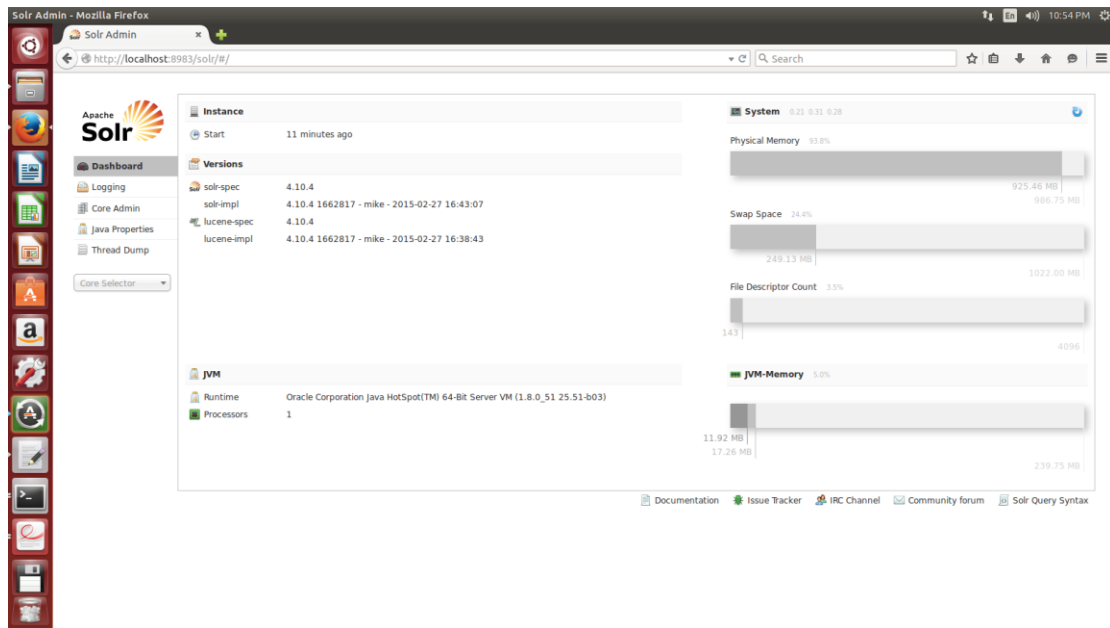
shift + 1 او shift + 2 لتغيير بين اللغتين

Close everything

### Step 6: ده خطوة التشغيل querying

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

This window will appear



Solr Admin - Mozilla Firefox

En10:56 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 2 days ago

Modified:

Num 68

Docs:

Max Doc: 68

Heap 9848

Memory

Usage:

Deleted 0

Docs:

Version: 5

Segment 2

Count:

Optimized: ✓

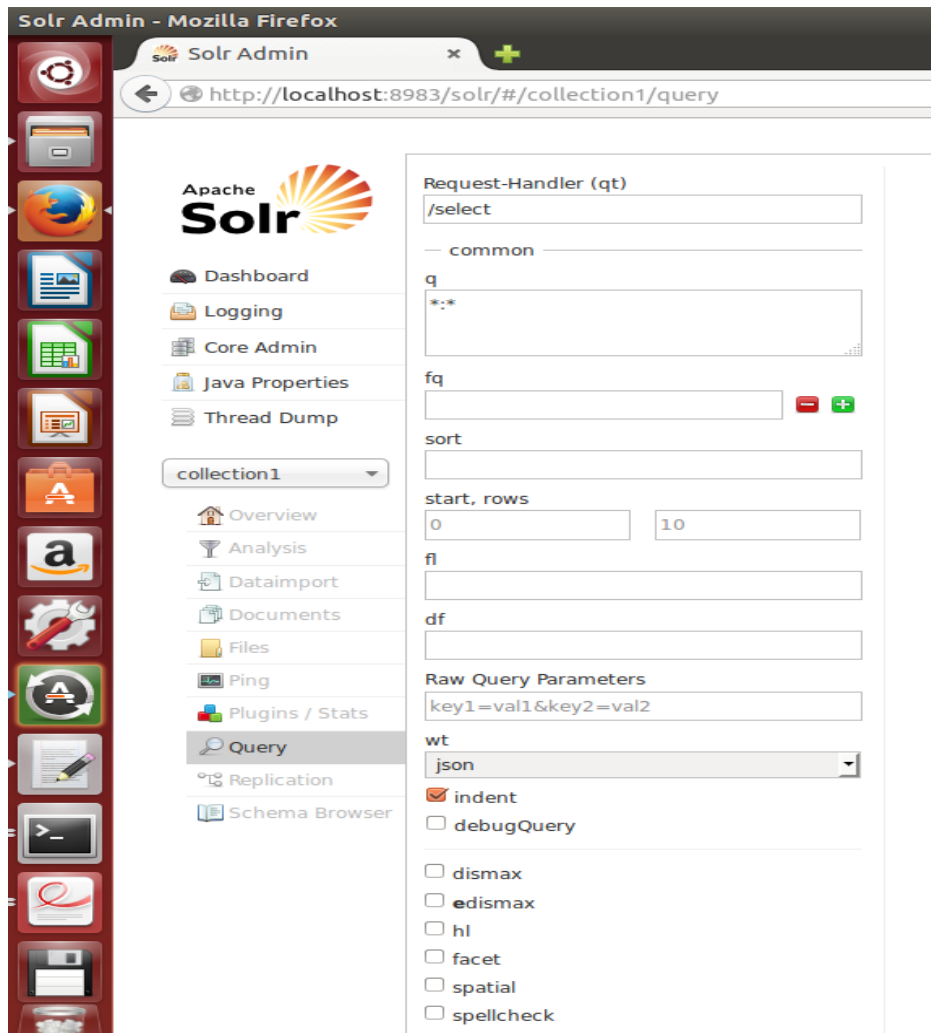
Current: ✓

Replication (Master)

Healthcheck

request handler is not configured a healthcheck file.

	Version	Gen	Ping	Size
Master (Searching)	1439062571846	3	677.98 KB	
Master (Replicable)	-	-	-	



Try a query

### **Step 8:** testing

Try words testing stemming and synonyms