

Installing IKE

Usage guide is good, but installation one isn't

Run Locally

1. Install PostgreSQL locally and create a database for use by IKE. IKE needs this to store the tables you create.
2. Modify the `Tablestore` key value setting in the [IKE config] (<https://github.com/allenai/ike/blob/master/src/main/resources/application.conf>) with appropriate database JDBC URL and credentials.
3. Run `sbt .`
4. Enter the `reStart` command.
5. Open <http://localhost:8080> in a browser.

Assumptions

- Ubuntu 18.04

Step 1

Installing Dependencies

Install SBT

- `sudo sbt about == 1.2.8`

Install Node

- Check versions
 - `sudo node -v == 11.12.0`
 - `sudo npm -v == 6.7.0`
- Use n module to manage different node version
 - `npm install -g n`
- The above versions worked for me, but other versions may work as well

Install PostgreSQL DB

- For more info, see link #2 under Sources slide

Install Java

- `java -version == 1.8`
- Tried java 10 and nothing worked

Step 2

Configuration

Create Postgres DB

- Create Postgres superuser: **ike**
- Create DB named **ike**
- Create Linux user named **ike**
- Set password to **ike**
 - Anything you want actually

Configure Java

- Set JAVA_HOME accordingly
 - `export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64`
- Put Postgres JDBC in correct place
 - Get jar from <https://jdbc.postgresql.org/download.html>
 - Move jar to `$JAVA_HOME/jre/lib/ext`

Add DB URL to IKE Config

- Config file in git repo
 - `src/main/resources/application.conf`
- Under Tablestore
 - Postgres JDBC Url
 - `jdbc:postgres://localhost:5432/ike`
 - Username: ike
 - Password: ike

Create Custom Index

- `ike/runMain org.allenai.ike.index.CreateIndex`
 - `-d /path/to/where/you/want/the/index`
 - **Note:** doesn't use `file://` prefix
 - `-t file:///path/to/your/data`
 - **Note:** has three slashes (`///`)
 - `-o`
 - **Note:** indicates that the data is a file with one document per line

Add Index URL to IKE Config

- Config file in git repo
 - `src/main/resources/application.conf`
- Under IkeToolWebapp
 - Make path equal the path to the newly created index

Warning: Corpus Size Must Be Small

- My hard (I)earned lesson
 - The text data of the case law of Arkansas is ~600 MB
 - The CreateIndex command creates an index of 12/13 times the size (takes too long)
 - Think of IKE's index as something like Solr's index
- Workaround
 - Split text data in ~100 MB files
 - Each file will take around 1.5 hours to transform into an index
 - Perks
 - Pipeline workflow by using IKE on data while waiting for the creation of another index
 - Parallelize index creation by running IKE on multiple computers

Step 3:

Run IKE

- Start local PostgreSQL server
 - `sudo systemctl start postgresql.service`
- Enter IKE git repo
- Run `sudo sbt`
- Run `reStart`
 - If you see “Connected Refused” errors, it’s fine
- Go to `localhost:8080`

Sources

1. <https://www.scala-sbt.org/1.x/docs/Installing-sbt-on-Linux.html>
2. <https://www.digitalocean.com/community/tutorials/how-to-install-and-use-postgresql-on-ubuntu-18-04>
3. <http://fm4dd.com/database/howto-install-PostgreSQL-jdbc.htm>
4. <https://jdbc.postgresql.org/documentation/80/connect.html>
5. <https://github.com/allenai/ike>
6. <https://github.com/tj/n>
7. Marc Matvienko