

# VIZLINC BUILD NOTES

## VIZLINC SOFTWARE

VizLinc is built out of the software in these repos:

<a href="#">vizlinc/vizlinc</a>	Gephi plugin; UI code. Non-maven project built in NetBeans
<a href="#">vizlinc/vizlincdb</a>	library used by both vizlinc and the ingester. It is built using maven because it depends on many third-party libraries, which are fetched automatically by maven.
<a href="#">vizlinc/ingester</a>	Groovy code and some data (NER model, geocoding data)

VizLinc is built using the latest NetBeans (7.4; 8.0 just came out and supports Java 8) and latest 64-bit Java 7: Java 7u51 as of this writing.

## VIZLINC INGESTION

The VizLinc ingester is composed of a number of Groovy classes, stored in the [vizlinc/ingester](#) git repository. The ingester also uses a number of classes from the [vizlinc/vizlincdb](#) package, which is also used by the main [vizlinc/vizlinc](#) code.

## RUNNING THE INGESTER

The GUI Ingester is documented in .docx and .pdf files that are checked into the git repository.

There is a command-line version of the ingester which can be invoked on Linux as:

```
$ cd ingester
$ ./run-groovy.sh src/Ingester.groovy -b basename -i input-docs -o out-dir -w work-dir [-m model]
```

<b>basename</b>	the prefix used for all the databases and other files, e.g. "isvg-2014-03-14"
<b>input-docs</b>	the top of a directory tree of documents from which to extract the text
<b>out-dir</b>	where the basename.h2, .lucene, and .graphml directories and files are written
<b>model</b>	the Stanford ner .ser.gz model file
<b>work-dir</b>	intermediate files are put here (.txt, .csv files, etc.)

Most of the individual steps of the pipeline are in separate .groovy files, and can be invoked individually as main programs for testing or running a single step of the ingestion process.

## RUN-GROOVY.SH AND RUN-GROOVY.BAT SCRIPTS

The **run-groovy.sh** script mentioned above has hardwired paths in it that refer to various pieces of software (groovy, gremlin, vizlincdb, and third-party jars) in **/home/userid/code** or similar places on a Windows machine. It also depends on the vizlincdb jar built in NetBeans, and points to that jar. There is a similar **run-groovy.bat** script for Windows.

## BUILDING AN EXECUTABLE FORM OF THE INGESTER

Download and install the following tools under Windows:

- Launch4j, <http://launch4j.sourceforge.net/>
- Apache Ant, <http://ant.apache.org/>
- Cygwin
- Netbeans and Maven Plugin

Modify the paths in the file **build.xml** to point to appropriate versions of Groovy, Gremlin, and VizLincDB. Note that vizlincdb must be built before building vizlinc-ingester.

```
$ ant          # compiles all groovy code into .class files
```

```
$ ant launch4j # above + builds the launch4j launcher for the GUI ingester (Windows only)
```

```
$ ant zip      # above + packages the GUI ingester, needed data and jars, and documentation into a zip file
```

## INGESTER PIPELINE

On the next page is a flow diagram of the current ingester pipeline. The top-level code running the pipeline is in **Ingestor.groovy**. **IngestorGUI.groovy** is a wrapper around it. There is a callback from **IngestorGUI** used by **Ingestor** to update the progress step indicator in the GUI.

