Datafile Builder README

Willie Rogers

January 13, 2023

Contents

1 Preliminaries: 1

2 Prerequisites 2

3 Getting the MetaMap Distribution 2

4 Extracting the distribution 2

5 Installing Data File Builder 2

6 Installing LVG on Mac OS/X 4

7 Instructions for using Data File Builder 4

8 Special Notes for Windows Users 5

1 Preliminaries:

This document is relevant only if

1. You do not want to use one of the data models provided with MetaMap, but want instead to build a custom Metathesaurus, and
2. You plan to build your custom Metathesaurus on Linux, or Mac OS/X.

The Data File Builder module, which constructs a custom Metathesaurus, requires certain GNU utilities, which are included in Linux distributions, some are not available on Mac OS/X. If you plan to use the Data File Builder on Mac OS/X, you need to download and install the GNU utilities.

It is *essential* that the GNU utilities be available, because the Data File Builder scripts use the GNU versions of programs such as grep, cut, join, sort, etc., and will not work properly if the Mac OS/X versions of some of these programs in /bin are used instead.

See README_macosx.html (http://metamap.nlm.nih.gov/README_macosx.html) for more information getting pre-compiled GNU tools binaries for Mac OS/X.

If the GNU utilities are already available on your system, you may skip to the Configuring the Shell Environment (§ **??**) section below.

2 Prerequisites

Platforms The MetaMap Datafile Builder currently only runs on the Linux and Mac OS/X.

Sun Java Runtime Environment (JRE) 1.6 or later Sun's Java 1.6.0 or later is required for use of the Datafile Builder Suite. Java is available from the "Developer Resources for Java Technology" website (http://java.sun.com/). **IMPORTANT**: If you have already installed MetaMap using Java 1.4 you must specify a Java 1.6 installation when re-running the MetaMap install program.

3 Getting the MetaMap Distribution

The public MetaMap distribution can be downloaded at the Download Section of the MetaMap website:

http://metamap.nlm.nih.gov/#Downloads

The latest version of this document can be downloaded at:

HTML http://metamap.nlm.nih.gov/README_dfb.html

PDF http://metamap.nlm.nih.gov/README_dfb.pdf

4 Extracting the distribution

Use the following `tar` command extract the distribution in the same directory where the Public MetaMap distribution was extracted:

```
% tar xvfj public_mm_linux_main_{year}.tar.bz2
% tar xvfj public_mm_linux_dfb__{year}.tar.bz2
```

If you don't have GNU tar installed use:

```
% bzip2 -dc public_mm_linux_main_{year}.tar.bz2 | tar xvf -
% bzip2 -dc public_mm_linux_dfb_{year}.tar.bz2 | tar xvf -
```

`Tar` will create the distribution directory `public_mm`. Note: The data compression program BZIP2 (available from http://www.bzip.org) is required to decompress the distributions. GNU tar is preferred, but not required to extract the contents of the distributions.

5 Installing Data File Builder

In addition to the following normal MetaMap install instructions, the Lexical Variant Generator (LVG) must be installed before running MetaMap's install program. LVG is part of the Lexical

Tools distribution and is available from the Lexical Systems Group (http://lexsrv3.nlm.nih.gov/Specialist/Summary/lexicalTools.html). If installing LVG on Mac OS/X or a platform other than Linux or Windows, use the instructions in the section on Installing LVG on Mac OS/X (§ 6).

Before using the MetaMap install program to install data file builder, the LVG bin directory ${LVG_DIR}/bin should be in the program path (LVG_DIR is where the LVG installation resides):

```
# in C Shell (csh or tcsh)
set path = ( $path <LVG_DIR>/bin )

# in Bourne Again Shell (bash)
export PATH=$PATH:<LVG_DIR>/bin

# Bourne Shell (sh)
PATH=$PATH:<LVG_DIR>/bin
export PATH
```

Connect to the new directory created by extracting the distribution and invoke the install program:

```
% cd <distribution directory>
% ./bin/install.sh
```

A sample run of the installation script follows:

```
Enter basedir of installation [/nfsvol/nlsaux15/public_mm] <user hits
                                                          return to get the default>
Basedir is set to /nfsvol/nlsaux15/public_mm.

The WSD Server requires Sun's Java Runtime Environment (JRE)
Sun's Java Developer Kit (JDK) will work as well. if the
command: "which" java returns /usr/local/jre1.4.2/bin/java, then the
JRE resides in /usr/local/jre1.4.2/.

Where does your distribution of Sun's JRE reside?
Enter home path of JRE (JDK) [/usr]: /nfsvol/nls/tools/Linux-i686/java1.4.2
Using /nfsvol/nls/tools/Linux-i686/java1.4.2 for JAVA_HOME.

/nfsvol/nlsaux15/public_mm/WSD_Server/config/disambServer.cfg generated
/nfsvol/nlsaux15/public_mm/WSD_Server/config/log4j.properties generated
/nfsvol/nlsaux15/public_mm/bin/SKRrun generated.
/nfsvol/nlsaux15/public_mm/bin/metamap07 generated.
/nfsvol/nlsaux15/public_mm/bin/wsdserverctl generated.
/nfsvol/nlsaux15/public_mm/bin/skrmedpostctl generated.
Install complete.
Would like to use a custom data set with MetaMap (use data file builder)? [yN]: <user types y a

running Data File Builder Install...
Is LVG installed? [yN] <The user types y and return>
```

```
running Data File Builder Install...
Enter home path of LVG [/nfsvol/nls/tools/Linux-i686/lvg{year}]: <user hits
                                                        return to get the default>

Using /nfsvol/nls/tools/Linux-i686/lvg{year} for LVG_DIR.

/nfsvol/nlsaux15/public_mm/scripts/dfbuilder/mm_variants/0doit.lvglab generated.
/nfsvol/nlsaux15/public_mm/scripts/dfbuilder/mm_variants/0doit.xwords generated.
Datafile Builder Setup is complete.
%
```

6 Installing LVG on Mac OS/X

If you are using a Mac there are special instructions for installing LVG on Mac OS/X:
Download the "Lite" version of Lvg for the year you wish to use (the example uses LVG 2010.) (see Lexical Tools Download Page: http://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/lvg/current/web/release/index.html)
extract it using tar and then for each of the files in the `lvg2010lite/bin` directory replace:

```
JAVA=/export/home/lu/Development/LVG/lvg2010/bin/jre1.6.0_14/bin/java
```

with:

```
JAVA=java
```

and:

```
LVG_DIR=/export/home/lu/Development/LVG/lvg2010
```

with:

```
LVG_DIR={where lvg is extracted}/lvg2010lite
```

Make sure all of the files in `lvg2010lite/bin` are executable:

```
chmod +x lvg2010lite/bin/*
```

The Java VM (java) is currently provided with Mac OS/X Snow Leopard.

7 Instructions for using Data File Builder

See the file datafilebuilder.pdf (Online version: http://metamap.nlm.nih.gov/datafilebuilder.pdf) for instructions on how to use datafile builder after installing it. (Windows Users please read next section on Special Notes for Windows Users (§ 8)).

8 Special Notes for Windows Users

Instead of running bin/install.sh after extracting the data file builder distribution Windows users should run either bin\ginstall.bat in the "Command Prompt" or "Windows PowerShell" window while in the "public_mm" directory or using Windows Explorer find the "public_mm" directory extracted from the distribution archive. Click on the icon with the name "Install MetaMap". The dialog will display an input box with the label "Location of the Public MetaMap Directory:" filled in with the probable location where you've installed the distribution.

Also note, when starting the MedPost/SKR tagger server (skrmedpostctl_start), it is best to use the Windows Command Prompt or Windows PowerShell in a separate window to start the server, rather than starting it from the BASH shell (public_mm/bin/sh.exe).

using the Command Prompt (or PowerShell):

```
> cd public_mm
> .\bin\skrmedpostctl_start.bat
... messages
```

After staring the Tagger server, in another Command Prompt (or PowerShell) window start the BASH shell in the `public_mm` directory:

```
> cd public_mm
> .\bin\sh
sh.exe"-3.1$ _
```

Then follow the instructions for using Data File Builder as mentioned in the previous section.