Datafile Builder README

Willie Rogers

November 2, 2009

Contents

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 Solaris.

The Data File Builder module, which constructs a custom Metathesaurus, requires certain GNU utilities, which are included in Linux distributions, but not with Solaris. If you plan to use the Data File Builder on Solaris, 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 Solaris versions of these programs in /bin are used instead.

The necessary GNU utilities may be freely downloaded and can be installed using the Solaris package manager. Knowledge of pkgadd and root privileges are required to install the GNU utilities; no C compiler is required, however.

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 Solaris 9 or greater.

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.

Solaris specific prerequisites See file README_solaris.html (online version: http://metamap. nlm.nih.gov/README_solaris.html) for more information on Solaris specific prerequisites.

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:

```
% bzip2 -dc public_mm_linux_2009.tar.bz2 | tar xvf -
% bzip2 -dc public_mm_linux_dfb_2009.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/Projects/lvg/current/index.html).

Before using the MetaMap install program to install data file builder, LVG's 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/lvg2009]: <user hits
                                                return to get the default>

Using /nfsvol/nls/tools/Linux-i686/lvg2009 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   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.