****

README FILE

Author: Jay Pisharath
Email: nu-minebench AT ece DOT northwestern DOT edu
Contents: Explains how to setup and execute NU-MineBench 2.0

****

TO DOWNLOAD:

Go to: http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html
Click on "Download" and download NU-MineBench-2.0.tar.gz

TO INSTALL:

```
tar -xvzf NU-MineBench-2.0.tar.gz
```

FILE DIRECTORY DESCRIPTION:

NU-MineBench is a collection of data mining applications. Currently there are 15 applications in the suite.

Here's the files structure. Lets call $DMHOME as the home for our data mining suite NU-MineBench-2.0. The following is the file organization within the root directory $DMHOME.

* **README.txt -** this file in txt format

* **commandLine.txt -** sample of the command line for executing each application.
Currently the sample commands use the default (most standard) parameters.
The command lines can similary be extended to other options/data sets.
This script uses $DMHOME as the home dir for applications.
So please "setenv" or "set" or "export" this variable before
trying out the commands.

* **src -** containing all the source files of the applications

* **datasets -** contains all the data sets for each application
 (subdirectory structure is the same as src)

Within src, the following subdirectories exist:

* **APR** - Apriori based association rule application (uses horizontal database)

* **ECLAT** - Another association rule application (uses vertical database)

* **Bayesian** - A naive bayesian classifier application

* **ScalParC** - A decision tree based classification application

* **birch -** Hierarchical clustering application

* **kmeans** - Partitioning based clustering application

* kmeans also contains a **fuzzy** based clustering application (execute with option -f to use fuzzy clustering)

* **hop** - Density based clustering application used in astrophysics

* **SNP** – Bayesian network based application for DNA sequence extraction

* **GeneNet** – Microarry based bayesian network application

* **semphy** – Phylogenetic tree based structure learning application

* **rsearch -** Stochastic Context-Free Grammer based RNA sequence search application

* **PLSA -** Dynamic programming based RNA/DNA sequence matching application

* **SVM-RFE -** Support Vector Machine based gene classification application

* **utility mining** – Utility based association rule application

****

Within datasets, you would find the relevant datasets for each application
(follows the same directory as above).

--
COMPILATION:
--

PLEASE TRY TO USE THE FOLLOWING CONFIGURATION OF COMPILERS

- GNU GCC/G++ version 3.2 or above
- Intel C++ Compiler version 7 or above
- Intel Fortran Compiler version 8 or above
- Intel Math Kernel Library 7.2 or above

APR:

```
cd $DMHOME/src/APR
make
```

ECLAT:

```
cd $DMHOME/src/ECLAT
make
```

Bayesian:

```
cd $DMHOME/src/Bayesian/bayes/src
make
```

ScalParC:

```
cd $DMHOME/src/ScalParC/
make
```

birch:

```
cd $DMHOME/src/birch
make
```

kmeans:

```
cd $DMHOME/src/kmeans
make example
```

hop:

```
cd $DMHOME/src/HOP
make
```

SNP:

```
cd $DMHOME/src/SNP/pnl.snps/pnl/c_pgmtk/src
make
cd $DMHOME/src/SNP/pnl.snps/snp
make
```

GeneNet:

```
cd $DMHOME/src/GeneNet/pnl.genenet/pnl_icc/c_pgmtk/src
make
cd $DMHOME/src/GeneNet/pnl.genenet/genenet
make
```

semphy:

```
cd $DMHOME/src/semphy
make -f Makefile.omp
```

For a complete serial (non-parallel) version of semphy:

```
cd $DMHOME/src/semphy
make
```

rsearch:
Offered in two parallel flavors, based on OpenMP and MPI.
For OpenMP version of rsearch,

```
cd $DMHOME/src/rsearch/rsearch-1.1.src-OpenMP
make
```

For the MPI version of rsearch,

```
cd $DMHOME/src/rsearch/rsearch-1.1.src-MPI
make
```

PLSA:

```
cd $DMHOME/src/PLSA
make -f Makefile.omp
```

SVM-RFE:

```
cd $DMHOME/src/SVM-RFE/svm_rfe
make -f makefile.omp
```

utility_mine:

```
$DMHOME/src/utility_mine/para_tran_utility
make
```

For a complete serial (non-parallel) version of utility_mine,

```
$DMHOME/src/utility_mine/tran_utility
make
```

EXECUTION:

See commandLine.txt file for sample

NOTE: for any application, just typing the application name without any command line options would list the actual command line options that are available to the user. Feel free to use all datasets provided.

Thanks for using NU-MineBench.
For issues send email to nu-minebench AT ece DOT northwestern DOT edu

****