MetaMap2016 Release Notes

François-Michel Lang

`metamap@nlm.nih.gov`

January 15, 2016

MetaMap2016 includes the much-requested and long-awaited functionality of ad-hoc exclusions of unwanted results, a new FAQ, and a utility program for converting UTF-8 files to their ASCII equivalents.

The Usage Notes for MetaMap2016 are available here.

Please let us know of any questions or comments about any of the following.

1   Ad-Hoc Exclusion of Unwanted Results

One of the most frequent feature requests received from users is the ability to exclude unwanted results from MetaMap output. Of course such exclusions could be implemented via postprocessing MetaMap output, but we have now provided this functionality through MetaMap directly.

Suppose a user considers these mappings infelicitous

- `BED` to C0596170 (preferred name `Binge eating disorder`) (removed from 2015AB)

- `PLAN` to C0270724 (preferred name `Infantile Neuroaxonal Dystrophy`)

and wishes to exclude them from MetaMap output. Simply create a plain text file

```
C0596170|BED
C0270724|PLAN
```

called, say, `Exclusions`, and call MetaMap as

`metamap --nomap Exclusions`

and the above mappings will be suppressed. The strings in the `Exclusions` file are case sensitive.

Some refinements:

- If the CUI is omitted, it is treated as a wildcard; all mappings to the string (regardless of the CUI) will be excluded.

- If the string is omitted, it is treated as a wildcard; all mappings to the CUI (regardless of the string) will be excluded.

2 System Upgrades and Changes

Past MetaMap releases have by default included a 32-bit binary. MetaMap2016, however, will default to a 64-bit binary. Please contact us if you need a 32-bit version of MetaMap. MetaMap downloads are available here.

We have also made available the MetaMap data files corresponding to the 2015AA and 2015AB UMLS releases. Dataset downloads are available here.

Other miscellaneous information:

- MetaMap2016 is statically and not dynamically linked, resulting in an appreciably larger file, but also simplifying the installation process.

- MetaMap2016 sends diagnostic output to stderr, and not stdout, as in previous versions.

- Finally, MetaMap2016 was upgraded to SICStus Prolog 4.3.2.

3 MetaMap Usage FAQ

We have created a MetaMap Usage FAQ, available here.

4 MetaMap UTF-8 Conversion Program

MetaMap accepts only ASCII input; users wishing to submit UTF-8 text to MetaMap will therefore need to convert their UTF-8 text to ASCII; in order to remove this burden from users, we have provided a simple UTF-8→ASCII conversion program, available here.

5 NLM Data Model now Includes All UMLS Data Sources

MetaMap introduced its three data models (Base, USAbase, and NLM) in the 2011 Release Notes. These three data models are designed to accommodate the UMLS source-vocabulary licensing permissions and processing requirements of as many users as possible. The USAbase model is MetaMap's default, but it can be overridden with `-V Base` or `-V NLM`. Beginning with MetaMap2016, the NLM data model now includes all UMLS sources.

MetaMap users are expected to adhere to all source-vocabulary copyright restrictions.

6 Changes to MetaMap Command-Line Options

We have made two changes to MetaMap command-line options:

1. The `--show_preferred_names_only` (`-O`) option has been removed.

2. The "-s" option, formerly called `--hide_semantic_types`, is now `--short_semantic_types`. Using this option will cause MetaMap to display in its Human-Readable Output e.g., `patf` and `inpo` instead of `Pathologic Function` and `Injury or Poisoning`. This option (in both its old and new versions) affects MetaMap's Human-Readable output only; invoking "-s" when generating any of Machine Output, XML Output, or Fielded MMI Output is an error.

7 Enhancement to MetaMap's Fielded MMI Output

MetaMap's Fielded MMI Output (generated via the `--fielded_mmi_output` (`-N`) option) contains positional information in its ninth pipe ("|") separated field. The format of the positional information has been updated to more accurately and unambiguously represent the position of the text.

Fielded MMI Output also now includes data about author-defined acronyms and abbreviations (AAs) that MetaMap found in the text, as well as about any user-defined AAs (UDAs). The use of UDAs is described here.

A full description of MetaMap2016's Fielded MMI Output is available here.

8 AAs' Pos Info Added to Machine Output and XML Output

MetaMap's Machine Output and XML Output now include information about the positional information of AAs.

8.1 Machine Output

MetaMap2014 Machine Output for AAs:

```
aas(["HA"*"heart attack"*(1,2,3,12)*[]]).
```

MetaMap2016 Machine Output for AAs, with changes shown in red; a bug preventing the display of an AA's CUI (if one is found) was also fixed.

```
aas(['HA'*'heart attack'*(1,2,3,12,14:2)*['C0027051']]).
```

8.2   XML Output

MetaMap2014 XML output for AAs:

```
<AAs Count="1">
 <AA>
  <AAText>HA</AAText>
  <AAExp>heart attack</AAExp>
  <AATokenNum>1</AATokenNum>
  <AALen>2</AALen>
  <AAExpTokenNum>3</AAExpTokenNum>
  <AAExpLen>12</AAExpLen>
  <AACUIs Count="0" />
 </AA>
</AAs>
```

MetaMap2016 XML output for AAs, with new XML elements shown in red; the bug preventing
the display of an AA's CUI (if one is found) was also fixed in XML output.

```
<AAs Count="1">
 <AA>
  <AAText>HA</AAText>
  <AAExp>heart attack</AAExp>
  <AATokenNum>1</AATokenNum>
  <AALen>2</AALen>
  <AAExpTokenNum>3</AAExpTokenNum>
  <AAExpLen>12</AAExpLen>
  <AAStartPos>14</AAStartPos>
  <AALen>2</AALen>
  <AACUIs Count="1">
   <AACUI>C0027051</AACUI>
  </AACUIs>
 </AA>
</AAs>
```