Fielded MetaMap Indexing (MMI) Output Explained (*Updated for MetaMap 2016 Output*)

The -N or --fielded_mmi_output option to **MetaMap** (https://metamap.nlm.nih.gov) displays a ranked list summary of all the mappings assigned to the text. Please see the ***MMI ranking function, 1997*** paper (https://ii.nlm.nih.gov/Publications/Papers/ranking.pdf) or ***MMI project description, 1997*** paper (https://ii.nlm.nih.gov/Publications/Papers/desc.pdf) for a detailed explanation of the MetaMap Indexing function. This document is designed to explain the information returned from MetaMap when using the option.
Please take note: Text highlighted in blue highlights MetaMap2014 changes from previous versions of the MMI output; changes in green show changes in MetaMap2016.

Sample Results:

```
24119710|MMI|637.30|Isopoda|C0598806|[euka]|["Isopod"-ab-1-"isopod"-adj-0,"Isopoda"-ti-1-"Isopoda"-noun-0]|TI;AB|228/6;136/7|B01.050.500.131.365.400
24119710|MMI|585.19|Subphylum Crustacea|C0010395|[euka]|["Crustacea"-ti-1-"Crustacea"-noun-0]|TI|125/9|B01.050.500.131.365
24119710|MMI|170.37|Effect|C1280500|[qlco]|["effects"-ti-1-"Effects"-noun-0]|TI|21/7|
24119710|MMI|3.44|Various patch test substance|C0440102|[irda]|["Various"-ab-1-"various"-adj-0]|AB|322/7|
 5538822|MMI|2.05|Basal Cell|C0596155|[cell]|["basal cells"-ab-41-"basal cells"-noun-0]|AB|7059/5,7073/5|
 5099124|MMI|2.51|Dominant|C1527180|[ftcn]|["Dominant"-ab-7-"dominant"-adj-0]|AB|[1351/8],[1437/8]|
23074497|MMI|7.59|Transplant Recipients|C0376387|[podg]|["Transplant Recipients"-ab-27-"recipients transplants"-noun-0]|AB|[4061/10,4075/11],
                                                                                                                [4166/10,4180/11]|M01.925
```

Bar "|" Separated Fields Included where Information is Available:

(1) **ID** – Unique identifier used to identify text being processed. If no identifier is found in the text, 00000000 will be displayed. For MEDLINE ASCII formatted input, the ID is considered the PMID.
(2) **MMI** – Always MMI. **Please Note**: This would previously have been shown as MM instead of MMI. *[Changed in MetaMap 2014]*
(3) **Score** – MetaMap Indexing (MMI) score with a maximum score of 1000.00. The higher the score, the greater the relevance of the UMLS concept according to the MMI algorithm. The MMI results are presented in highest-to-lowest relevance order.
(4) **UMLS Concept Preferred Name** – The UMLS preferred name for the UMLS concept identified in the text.
(5) **UMLS Concept Unique Identifier (CUI)** – The CUI for the identified UMLS concept.
(6) **Semantic Type List** – Comma-separated list of Semantic Type abbreviations for the identified UMLS concept. More information on ***Semantic Types and Semantic Groups*** can be found at (https://metamap.nlm.nih.gov/SemanticTypesAndGroups.shtml).
(7) **Trigger Information** – Comma separated sextuple showing what triggered MMI to identify this UMLS concept. Ordering tends to go from last to first found.
The sextuple consists of "UMLS concept-loc-locPos-text-Part of Speech-Negation Flag".
- **UMLS Concept (Preferred or Synonym Text)**
- **loc** – Location in the text if identifiable. ti – Title, ab – Abstract, and tx – Free Text
- **locPos** – Number of the utterance within the loc starting with one (1). For example, "ti-1" denotes the first utterance in the title.
- **text** – The actual text mapped to this UMLS concept identification.
- **Part of Speech** – Determined by the MedPost Tagger or Lexical Lookup *[New field in MetaMap 2014]*
- **Negation Flag** – 1 if the text is considered Negated by MetaMap and 0 if it is not considered Negated. *[New field in MetaMap 2014]*
(8) **Location** – Summarizes where UMLS concept was found. TI – Title, AB – Abstract, TX – Free Text, TI;AB – Title and Abstract
(9) **Positional Information** – Semicolon-separated list of positional-information terns, showing StartPos, slash (/), and Length of each trigger identified in the Trigger Information field. The Positional Information field will contain as many semicolon-separated chunks of positional information as there are sextuples of trigger information in the Trigger Information field. StartPos begins at position zero (0) of the input text. Each semicolon-separated chunk of positional information can take one of four forms:
   a) The simplest form: StartPos/Length, e.g., 228/6, 136/7, 125/9, 21/7, 322/7 in the first four examples (all PMID 24119710), meaning that the concept in question was identified in a single span in the text beginning at position 228 and spanning 6 characters, or beginning at position 136 and spanning 7 characters, etc. Note that the first example contains two sextuples in the Trigger Information field, and therefore two semicolon-separated chunks of positional information.
   b) Multiple comma-separated StartPos/Length pairs, e.g., 7059/5,7073/5 in the fifth example (PMID 5538822), indicating disjoint text strings mapped to one concept; an example is "basal" and "cells" in the phrase "basal layer cells".

Fielded MetaMap Indexing (MMI) Output Explained (*Updated for MetaMap 2016 Output*)

    c)   Multiple comma-separated *bracketed* StartPos/Length pairs, e.g., [1351/8],[1437/8] in the next-to-last example (PMID 5099124) , indicating multiple occurrences of the same concept in the same utterance. Note the semantic difference between 7059/5,7073/5 and [1351/8],[1437/8]: The former indicates a single mapping to a concept from text consisting of disjoint text spans, and the latter, multiple occurrences of a concept in an utterance. In the previous version of the positional-information notation, these two semantically different terms were represented identically.

    d)   Finally, forms (b) and (c) above can in rare cases be combined, e.g., [4061/10,4075/11],[4166/10,4180/11] in the last example (PMID 23074497) . This last form signals multiple occurrences of a concept mapping in an utterance, each of which consists of disjoint text spans. *[Changed in MetaMap 2016]*

(10) **Treecode(s)** – Semicolon-separated list of any MeSH treecode(s) (https://www.nlm.nih.gov/mesh/meshhome.html) associated with the UMLS concept; field may be null if no treecodes were found for the concept.

Fielded MMI Output also includes information about acronyms and abbreviations (AAs) that MetaMap identified in the text. Sample AA lines in MMI output are shown below; the positional information field is new to MetaMap2016.:

```
23074487|AA|FY|fiscal years|1|2|3|12|9362:2
23074399|AA|DORs|diagnostic odds ratios|1|4|5|22|8926:4
17342196|AA|PCBs|polychlorinated biphenyls|1|4|3|25|2304:4
23074488|AA|ART|assisted reproductive technologies|1|3|5|34|1197:3
12952502|AA|VRBPAC|Vaccination and Related Biologicals Products Advisory Committee|1|6|13|63|9946:6
23074402|AA|THETA|Toronto Health Economics and Technology Assessment Collaborative|1|5|13|64|1291:5
23074475|AA|INAHTA|International Network of Agencies for Health Technology Assessments|1|6|15|67|8780:6
```

AA lines consist of nine pipe-separated fields:

(1) **ID** – Unique identifier used to identify text being processed. If no identifier is found in the text, 00000000 will be displayed. For MEDLINE ASCII formatted input, the ID is considered the PMID.

(2) **AA –** Always "AA", to differentiate AA lines from the MMI lines described above.

(3) **Short form** – The short form of the acronym/abbreviation.

(4) **Long form** – The long form or expansion of the acronym/abbreviation.

(5) **# of tokens in short form** – The number of tokens (including whitespace tokens) in the short form.

(6) **# of characters in short form** – The number of characters in the short form.

(7) **# of tokens in long form** – The number of tokens (including whitespace tokens) in the long form.

(8) **# of characters in long form** – The number of characters in the long form.

(9) **Positional information of short form** – The starting position of the short form followed by a colon (":") followed by the character length of the short form.

Finally, Fielded MMI Output now also includes information about user-defined acronyms and abbreviations (UDAs) found in the text, which is presented as follows:

```
00000000|UA|CHOP|Cyclophosphamide, Hydroxydaunomycin, Oncovin & Prednisolone|1|4|11|59|8:4
```

The nine fields are identical to those described above other than the second, which appears as "UA" instead of "AA".

More information about UDAs is available in Section 9 of the MetaMap 2011 Release Notes (https://metamap.nlm.nih.gov/Docs/MM_2011_ReleaseNotes.pdf) .