

Fielded MetaMap Indexing (MMI) Output Explained (*Updated for MetaMap 2014 Output*)

The `-N` or `--fielded_mmi_output` option to **MetaMap** (<http://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 (<http://ii.nlm.nih.gov/Publications/Papers/ranking.pdf>) or *MMI project description, 1997* paper (<http://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: Everything highlighted in **blue** showcases changes from previous versions of the MMI output.

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|14.64|Cadmium|C0006632|[elii,hops]|["Cadmium"-ab-1-"cadmium"-noun-0]|AB|359:7|D01.268.556.137;D01.268.956.061;D01.552.544.137;x.x.x.x
24119710|MMI|10.42|Length|C1444754|[qncq]|["Length"-ab-4-"length"-noun-0]|AB|917:6^1049:6|
24119710|MMI|3.44|Various patch test substance|C0440102|[irda]|["Various"-ab-1-"various"-adj-0]|AB|322:7|
```

Bar “|” Separated Fields Included where Information is Available:

- ❖ **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.
- ❖ **MMI** – Always MMI. **Please Note:** This would previously have been shown as MM instead of MMI. **[Changed in MetaMap 2014]**
- ❖ **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.
- ❖ **UMLS Concept Preferred Name** – The UMLS concept’s preferred name for the UMLS concept identified in the text.
- ❖ **UMLS Concept Unique Identifier (CUI)** – The CUI for the identified UMLS concept.
- ❖ **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 (<http://metamap.nlm.nih.gov/SemanticTypesAndGroups.shtml>).
- ❖ **Trigger Information** – Comma separated sextuple showing what triggered MMI to identify this UMLS concept. Ordering tends to go 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 first utterance in title.
 - **text** – The actual text mapped to this UMLS concept identification.
 - **Part of Speech** – Determined by the MedPost Tagger or Lexicon 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]**
- ❖ **Location** – Summarizes where UMLS concept was found. TI – Title, AB – Abstract, TX – Free Text, TI;AB – Title and Abstract
- ❖ **Positional Information** – Semi-colon separated list of positional information doubles showing StartPos, colon (:), and Length of each trigger identified in the Trigger Information field. StartPos begins at position zero (0) of the input text. For example, in the first line of sample results above, we have “[228:6;136:7]” positional information showing we found two triggers in the text. One starting at position 228 which is 6 characters in length (*isopod*) and the second at position 136 consisting of 7 characters (*Isopoda*). The positional information like the trigger information tends to go from last found to first found. If there are multiple occurrences in the same utterance as in “Length” above “[917:6^1049:6]”, the occurrences are separated by an up caret character “^”. **[Changed in MetaMap 2014]**
- ❖ **Treecode(s)** – Semicolon-separated list of any MeSH treecode(s) (<http://www.nlm.nih.gov/mesh/meshhome.html>) associated with the UMLS concept; field may be null if no treecodes were found for the concept.