

- ⇒ [MetaMap Home](#)
- ⇒ [FAQ Home](#)
- ⇒ [Help/Questions](#)

Using MetaMap to Process Clinical Text

As noted elsewhere in these FAQ, MetaMap was originally developed to process articles in the biomedical literature. However, several features were added specifically to assist in the processing of clinical text.

- **Negation Detection:** Knowing if a concept is negated is far more important in analyzing clinical text than it is for biomedical literature. We recommend invoking MetaMap with `--negex` to display information about negated concepts in [Human-Readable Output](#) output (the `--negex` option is redundant when generating [XML Output](#), [Prolog Machine Output](#), or [Fielded MMI Output](#), because these three output formats automatically include negation information). MetaMap's negation detection is based on Wendy Chapman's [NegEx](#), and includes some significant extensions, documented [here](#).
- **BlankLines:** MetaMap by default treats any number of consecutive blank lines as an input-record separator. However, this default behavior is often undesirable for analyzing clinical text, because clinical notes often include multiple blank or whitespace lines in the middle of reports. Invoking MetaMap with `--blanklines N` (where N is be a positive integer), will result in MetaMap's not ending an input document until N blank/whitespace lines have been read, thereby allowing an entire clinical note to be analyzed as a single document. Specifying N as zero will cause MetaMap to detect no input-record separators at all; consequently, the entire input file will be treated as a single document.
- **User-Defined Acronyms:** Clinical text is full of idiosyncratic, often site-specific acronyms that are neither defined in the text nor included in the UMLS, and can therefore not be identified by MetaMap. User-defined acronyms (UDAs) allow MetaMap users to supply their own definitions for such acronyms, thereby allowing MetaMap to map them to UMLS terms.

Simply create a file (let's call it `UDAfile`) containing lines such as

```
PET|Positron Emitting Tomography
CAT|Computerized Axial Tomography
DRSP|drug-resistant streptococcus pneumoniae
NIDR|National Infectious Disease Register
```

and invoke MetaMap with `--UDA UDAfile`. MetaMap will then expand all occurrences of the user-defined acronyms defined in `UDAfile` that are found in the input text. UDAs are case sensitive!

The last line of the UDA file must end with a newline character!

Far more detailed information about UDAs can be found in Section 9 of [here](#).

One small change for MetaMap2020 should be noted; this statement no longer holds:

UPDATED for MetaMap2020

MetaMap will consider the shorter string to be the AA, and the longer one the expansion, regardless of which appears first in a given line.

Beginning with MetaMap2020, the text before the pipe symbol will be replaced by the text after the pipe symbol, regardless of the length of the two text segments. This change is intended to allow users to replace any string with any other.