MetaMap's Default Input:
Unformatted English Free Text   UPDATED for MetaMap2020

MetaMap's default input is **unformatted English free text**.

Beginning with MetaMap2020, MetaMap will accept UTF-8 input. UPDATED for MetaMap2020

Moreover, beginning with the 2020AA release of the UMLS, MetaMap's datafiles will be UTF-8. UPDATED for MetaMap2020

Input records are by default separated by a blank line; however see the other sections in FAQ #2 for additional input formats.

If needed, a Java program that converts the most common non-ASCII characters to their closest ASCII equivalents is available here.

The last line of all MetaMap input files must end with a newline character!

Because MetaMap by default processes English text only, MetaMap's UMLS-based datafiles that are used to map text to UMLS concepts include only the English strings from the UMLS sources (i.e., `LAT == "ENG"`). Users needing to include sources in languages other than English in order to process non-English text should create their own UMLS-like database using our Data File Builder suite, available here.