Suppressing Numerical Concepts with \texttt{--no_nums}

The ability to suppress numerical concepts was introduced in MetaMap16V2; previous MetaMap versions do not include this functionality.

We have found that many mostly-numerical strings in certain semantic types (e.g., geoa, fndg, spco, qnco, tmco) add little value to biomedical named-entity recognition.

Some typical numerical strings that we consider too noisy to be of much use are

\begin{verbatim}
0.45 80% 29+ 3/4 1 - 9% 1:100 >97 0.010``
\end{verbatim}

The option \texttt{--no_nums} option can be used to suppress such UMLS concepts in which at least 50\% of the non-whitespace characters are digits. This option requires specifying a comma-separated list of Semantic Types as an option, e.g.,

\begin{verbatim}
--no_nums fndg
--no_nums fndg,spco
\end{verbatim}

Finally, to exclude all such mostly-numerical concepts, use the pseudo-SemType \texttt{all} or \texttt{ALL}:

\begin{verbatim}
--no_nums all
--no_nums ALL
\end{verbatim}