

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

## Processing Tabular Data with `--num_break`

Processing Tabular Data was introduced in MetaMap16V2; previous MetaMap versions do not include this functionality.

This page provides a technique for solving a problem similar to that presented in the [bulleted lists](#) page.

As explained [here](#), MetaMap accepts only ASCII input. Consequently, MetaMap users wishing to analyze the text of web pages will first have to screen-scrape the display or otherwise extract the text from the markup language. If a web page contains a table such as

<b>Blood and lymphatic system disorders</b>				
Neutropenia	52.8	48.9	49.6	45.8
Anemia	23.1	2.5	18.9	3.5
Leukopenia	18.2	12.3	20.4	14.6
Febrile neutropenia <sup>2</sup>	13.8	13.0	7.6	7.3
<b>Nervous system disorders</b>				
Neuropathy peripheral	32.4	3.2	33.8	2.0
Headache	20.9	1.2	16.9	0.5
Dysgeusia	18.4	0.0	15.6	0.0
Dizziness	12.5	0.5	12.1	0.0

the extracted text will probably look something like the following:

```
Blood and lymphatic system disorders
Neutropenia      52.8      48.9      49.6      45.8
Anemia           23.1       2.5      18.9       3.5
Leukopenia       18.2      12.3      20.4      14.6
Febrile neutropenia 13.8      13.0       7.6       7.3
Nervous system disorders
Neuropathy peripheral 32.4      3.2      33.8       2.0
Headache         20.9       1.2      16.9       0.5
Dysgeusia        18.4       0.0      15.6       0.0
Dizziness        12.5       0.5      12.1       0.0
```

which is essentially a noun pile with many interspersed numbers. MetaMap's phrase chunker will detect no syntactic structure in the text, and consequently analyze the text of the entire table as one phrase, which will require extensive processing time.

To solve this problem, we recommend either [Suppressing Numerical Concepts](#), or, better yet, using the `--num_break` option, which will force the phrase chunker to declare a phrase break after encountering two consecutive numbers.

With the `--num_break` option, the smaller phrases analyzed, displayed one per line, will be the following, saving considerable processing time with no degradation of results:

```
Blood
and
lymphatic system disorders Neutropenia      52.8
48.9      49.6
```

45.8	Anemia	23.1	
2.5	18.9		
3.5	Leukopenia	18.2	
12.3	20.4		
14.6	Febrile neutropenia	13.8	
13.0	7.6		
7.3	Nervous system disorders	Neuropathy peripheral	32.4
3.2	33.8		
2.0	Headache	20.9	
1.2	16.9		
0.5	Dysgeusia	18.4	
0.0	15.6		
0.0	Dizziness	12.5	
0.5	12.1		
0.0			