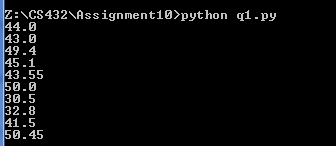
1. For the first question I used a code from <https://github.com/rcond002/cs532-s16/blob/master/Assignment10/Question1/A10Q1.py> named q1.py and modified it to work with my input blogdata1.txt. The results are in a screenshot here:



1. I did not do this.
2. For this one I used two scripts made by Gabriel Marquez called twitter\_part2.py and twitter\_part3.py to obtain the results I needed. Here are the original table and the new one for comparison:

Original:

|  |  |
| --- | --- |
| Mementos/URL | Frequency |
| 0 | 958 |
| 1 | 6 |
| 2 | 4 |
| 3 | 4 |
| 4 | 1 |
| 5 | 1 |
| 9 | 2 |
| 12 | 1 |
| 13 | 1 |
| 15 | 1 |
| 16 | 7 |
| 17 | 1 |
| 19 | 2 |
| 35 | 1 |
| 40 | 1 |
| 41 | 1 |
| 44 | 2 |
| 45 | 1 |
| 50 | 1 |
| 71 | 1 |
| 160 | 1 |
| 218 | 1 |
| 297 | 1 |
| 1105 | 1 |

New:

|  |  |
| --- | --- |
| Mementos/URL | Frequency |
| 0 | 958 |
| 1 | 5 |
| 2 | 4 |
| 3 | 4 |
| 5 | 2 |
| 9 | 2 |
| 12 | 1 |
| 13 | 1 |
| 15 | 1 |
| 16 | 7 |
| 17 | 1 |
| 19 | 2 |
| 35 | 1 |
| 40 | 1 |
| 41 | 1 |
| 44 | 2 |
| 45 | 1 |
| 50 | 1 |
| 71 | 1 |
| 160 | 1 |
| 253 | 1 |
| 297 | 1 |
| 1139 | 1 |

Difference Table:

