1. The original 1000 URI’s html and text documents that were extracted are in the “Content” directory. The URI’s whose content I used are in the “searchGoogle” directory. I used “Google” as my search term.

2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Doc # | TF | IDF | TF-IDF | URI |
| 1 | 4/865=0.0046 | 1.9469 | 0.009003 | [http://creative-punch.net](http://creative-punch.net/) |
| 2 | 5/1762=0.0028 | 1.9469 | 0.005525 | <http://www.alephnaught.com/Blog/> |
| 3 | 3/2222=0.0014 | 1.9469 | 0.002629 | <http://cssmatter.com/> |
| 4 | 5/912=0.0055 | 1.9469 | 0.010674 | <https://www.youtube.com/channel/UCkA0X11C61VYxaijEilJWzw> |
| 5 | 19/291=0.0653 | 1.9469 | 0.127117 | <http://bookterabyte.hateblo.jp/> |
| 6 | 2/255=0.0078 | 1.9469 | 0.01527 | [http://cha1tanya.com](http://cha1tanya.com/) |
| 7 | 2/300=0.0067 | 1.9469 | 0.012979 | [http://daanlenaerts.com](http://daanlenaerts.com/) |
| 8 | 1/3889=0.0003 | 1.9469 | 0.000501 | [http://hereisyourwaytoescape.tumblr.com](http://hereisyourwaytoescape.tumblr.com/) |
| 9 | 4/2100=0.0019 | 1.9469 | 0.003708 | [http://jobnec.com](http://jobnec.com/) |
| 10 | 1/1430=0.0007 | 1.9469 | 0.001361 | [http://crim5onviolet.tumblr.com](http://crim5onviolet.tumblr.com/) |

3.

|  |  |  |
| --- | --- | --- |
| Doc # | Page Rank | URI |
| 2 | 0.4 | <http://www.alephnaught.com/Blog/> |
| 3 | 0.2 | <http://cssmatter.com/> |
| 6 | 0.2 | [http://cha1tanya.com](http://cha1tanya.com/) |
| 10 | 0.1 | [http://crim5onviolet.tumblr.com](http://crim5onviolet.tumblr.com/) |
| 1 | 0 | [http://creative-punch.net](http://creative-punch.net/) |
| 4 | 0 | <https://www.youtube.com/channel/UCkA0X11C61VYxaijEilJWzw> |
| 5 | 0 | <http://bookterabyte.hateblo.jp/> |
| 7 | 0 | [http://daanlenaerts.com](http://daanlenaerts.com/) |
| 8 | 0 | [http://hereisyourwaytoescape.tumblr.com](http://hereisyourwaytoescape.tumblr.com/) |
| 9 | 0 | [http://jobnec.com](http://jobnec.com/) |

4. Kendall tau=-0.083 p-value=0.8379