Gorilla–Sea Cucumber Hash Report

Results

The following table gives the similarity between each pair of species as a number between 0 and 1, higher values meaning “more similar.”
We have used the hash function `h(S)=String.hashCode()%d` with d = 100 and kgrams of length k = 5.
The results emerged from the analysis of file `HbB_FASTAs-in.txt.`
As can be seen, the species closest to us is the Gorilla.

	Human	Gorilla	Monkey	Horse	Deer	Pig	Cow	Gull	Trout	R. Cod	Lamprey	Sea C.
Human	1	0.986	0.876	0.791	0.727	0.732	0.816	0.646	0.653	0.562	0.562	0.534
Gorilla	0.98	1	0.852	0.791	0.711	0.694	0.838	0.633	0.628	0.564	0.578	0.515
Monkey	0.876	0.852	1	0.788	0.726	0.780	0.787	0.656	0.661	0.523	0.536	0.567
Horse	0.791	0.791	0.788	1	0.680	0.690	0.785	0.597	0.622	0.597	0.599	0.592
Deer	0.727	0.711	0.726	0.680	1	0.741	0.744	0.647	0.632	0.581	0.578	0.583
Pig	0.732	0.694	0.780	0.690	0.741	1	0.700	0.616	0.648	0.521	0.577	0.559
Cow	0.816	0.838	0.787	0.785	0.747	0.700	1	0.627	0.609	0.552	0.564	0.520
Gull	0.646	0.633	0.656	0.597	0.647	0.617	0.627	1	0.668	0.564	0.618	0.576
Trout	0.653	0.628	0.661	0.622	0.632	0.648	0.609	0.668	1	0.608	0.610	0.660
R. Cod	0.562	0.564	0.523	0.597	0.581	0.521	0.522	0.564	0.608	1	0.625	0.676
Lamprey	0.562	0.578	0.536	0.599	0.578	0.577	0.764	0.618	0.610	0.625	1	0.664
Sea C.	0.534	0.515	0.567	0.592	0.583	0.559	0.520	0.576	0.660	0.676	0.664	1

Our `comparator(int creatureTypeIndex)` computes dot product, magnitudes and finally similarity of two different species. It was tested directly on the file at hand, using the `Creature` object `profile` field and giving for example the following profile/vector results (100-dimensional):

Human profile:{1=1, 2=2, 3=2, 4=1, 5=2, 6=2, 7=4, 8=1, 10=2, 12=1, 13=1, 14=2, 15=2, 17=2, 16=1, 19=1, 18=3, 21=2, 23=1, 25=1, 24=1, 27=3, 26=2, 28=4, 31=2, 30=1, 34=2, 35=1, 33=3, 38=1, 39=1, 36=1, 37=1, 42=1, 43=1, 40=2, 41=4, 46=1, 47=2, 51=1, 50=1, 48=5, 55=1, 54=1, 53=2, 52=2, 59=1, 58=6, 57=2, 56=1, 63=1, 62=2, 60=1, 70=2, 71=1, 64=2, 65=1, 66=2, 67=3, 76=3, 77=3, 78=1, 79=2, 72=2, 73=3, 75=1, 84=2, 87=1, 86=1, 80=2, 83=4, 82=2, 93=2, 92=1, 95=2, 91=5, 98=1, 99=1}

Gorilla profile:{1=1, 2=2, 3=2, 4=1, 5=2, 6=2, 7=4, 8=1, 10=2, 12=1, 13=1, 14=2, 15=2, 17=2, 16=1, 19=1, 18=3, 21=2, 23=1, 25=1, 24=1, 27=3, 26=2, 28=4, 31=2, 30=1, 34=2, 35=1, 33=3, 38=1, 39=1, 36=1, 37=1, 42=1, 43=1, 40=2, 41=4, 46=1, 47=2, 51=1, 50=1, 48=5, 55=1, 54=1, 53=2, 52=2, 59=1, 58=6, 57=2, 56=1, 63=1, 62=2, 60=1, 70=2, 71=1, 64=2, 65=1, 66=2, 67=3, 76=3, 77=3, 78=1, 79=2, 72=2, 73=3, 75=1, 84=2, 87=1, 86=1, 80=2, 83=4, 82=2, 93=2, 92=1, 95=2, 91=5, 98=1, 99=1}

This Java implementation took 1,839 seconds to compute the similarity of each of the 12 species.