Gorilla–Sea Cucumber Hash

Dennis Thinh Tan, William Diedrichsen Marstrand

April 9, 2015

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

```
private static int hash(String string)
  {
    return (string.hashCode()) % d;
  }
```

with $d = 10000$ and k-grams of length $k = 10$. 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 Cuc.
Human	1	0.948	0.842	0.729	0.910	0.556	0.513	0.465	0.412	0.364	0.328	0.294
Gorilla	0.948	1	0.761	0.606	0.488	0.443	0.424	0.375	0.328	0.285	0.255	0.227
Monkey	0.842	0.761	1	0.654	0.480	0.412	0.384	0.348	0.295	0.256	0.227	0.198
Horse	0.729	0.606	0.654	1	0.553	0.439	0.393	0.340	0.281	0.235	0.208	0.179
Deer	0.610	0.488	0.480	0.553	1	0.558	0.411	0.311	0.247	0.200	0.171	0.144
Pig	0.556	0.443	0.412	0.439	0.558	1	0.587	0.407	0.304	0.236	0.196	0.163
Cow	0.513	0.424	0.384	0.393	0.411	0.407	1	0.555	0.367	0.272	0.215	0.173
Gull	0.465	0.375	0.348	0.340	0.311	0.407	0.555	1	0.482	0.305	0.225	0.173
Trout	0.412	0. 328	0.295	0.281	0.247	0.304	0.367	0.482	1	0.472	0.314	0.225
Rockcod	0.364	0.285	0.256	0.235	0.200	0.236	0.272	0.305	0.472	1	0.511	0.324
Lamprey	0.328	0.255	0.227	0.208	0.171	0.196	0.215	0.226	0.314	0.511	1	0.460
Sea-Cucumber	0.294	0.227	0.198	0.179	0.144	0.163	0.173	0.173	0.225	0.324	0.460	1

Tests

Our static method `double cos_angle(int[] p, int[] q)` computes the cosine of the angle of two vectors of the same length d. We have tested it on the following examples:

p	q	d	value returned	length of q
$(0,1)$	$(0,1)$	2	1	1
$(0,1)$	$(0,2)$	2	1	2
$(0,1)$	$(1,0)$	2	0.0	1
$(0,1)$	$(0,-1)$	2	-1	1
$(0,\ldots,0,1)$	$(1,0,\ldots,0)$	1000	0	1