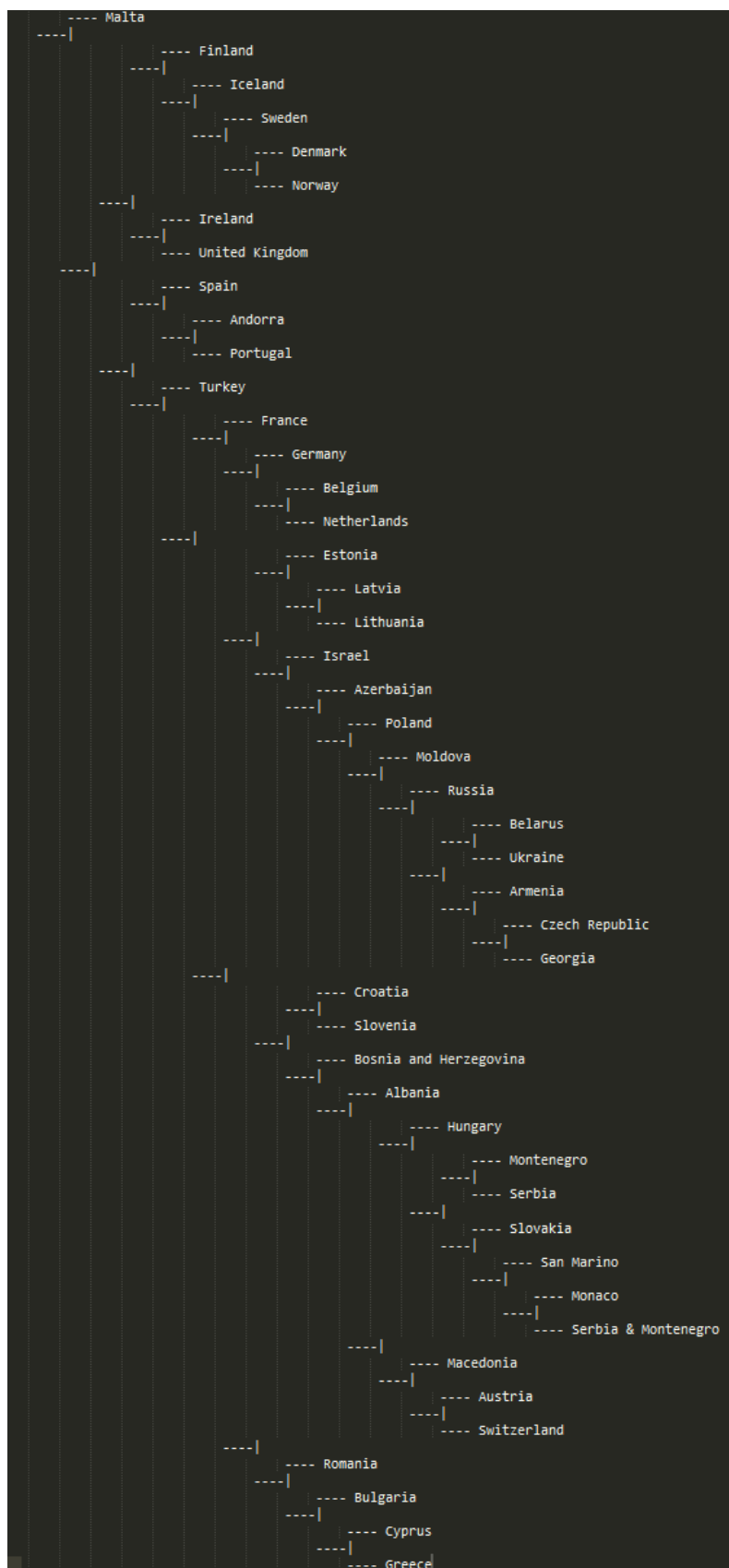


**Podatki.** Podatki so podani v obliki tabele, ki je vsebovala države, njihove pesmi in glasovanja na finalnem delu Evrovizije v razponu 11 let. Tabela je vsebovala nekaj atributov, ki niso relevantni in so bili zato odstranjeni (ime pesmi, izvajalec in drugo). Uporabljene so bile ocene ostalih držav za določeno pesem druge države. Število držav, torej atributov, je 47, število primerov pa je 291. Zaloga vrednosti glasov je med 0 in 12, povprečno število glasov pa je 2,5. Vsako leto se niso vse države uvrstile v finalni del, zato manjka 34% podatkov. Za posamezno državo dobimo profil glasovanja kar kot posamezen atribut države. Ta vsebuje vse primere kako je država glasovala za ostale. Če izračunamo razdaljo med temi profili, dobimo blizu skupaj države, ki so podobno glasovale.

**Računanje razdalj.** Razdalje med profili so bile izračunane z Evklidsko razdaljo, ki računa razlike med glasovanji na podlagi posameznih primerov znotraj dveh profilov. Na podlagi teh razdalj je bila ustvarjena prva gruča. Za računanje razdale med njimi je bila uporabljena povprečna Evklidska razdalja med gručami. Torej se je izračunala razdalja med vsemi elementi obeh gruč in se nato delila s številom povezav.

V tabeli je primanjkovalo kar 34% podatkov. Če ne bi bili umetno dodani, bi izgubili velik del podatkov. Zato je bilo za manjkajoče primere uporabljeno povprečje glasovanj iz ostalih let. Torej če v določenem letu neka država ni glasovala za drugo državo, se je ta glas nadomestil s povprečjem iz glasovanj ostalih let za par država glasovalka - glasovana država.

**Dendogram** Prikazan je na sliki 1, v prilogi B pa je za boljo preglednost podana še tekstovna datoteka dendograma.



Slika 1: Tekstovni dendrogram podobnosti profilov glasovanja držav.

**Skupine in njihove preferenčne izbire.** Skupine so bile oblikovane vizualno glede na vertikalni presek dendograma (razen Malte in Turčije, ki sta izstopali in sta bili dodani v zadnjo skupino). Preferiti so bili izbrani tako, da je bila vzeta država, ki je od ostalih držav dobila maksimalno število glasov.

Tabela 1: Glasovalne skupine in njihovi preferiti .

skupina	države	preferiti
1	Finland, Iceland, Sweden, Denmark, Norway	Sweden
2	Ireland, United Kingdom	United Kingdom
3	Spain, Andorra, Portugal	Spain
4	France, Germany, Belgium, Netherlands	Germany
5	Estonia, Latvia, Lithuania	Lithuania
6	Russia, Belarus, Ukraine, Armenia, Czech Republic, Moldova, Georgia, Poland	Russia Poland
7	Azerbaijan, Israel	Azerbaijan
8	Romania, Bulgaria, Cyprus, Greece	Greece
9	Croatia, Slovenia, Bosnia and Herzegovina	Croatia
10	Albania, Hungary, Montenegro, Serbia, Slovakia San Marino, Monaco, Serbia & Montenegro	Serbia
11	Macedonia, Austria, Switzerland	Switzerland
12	Malta, Turkey	

**Izjava o izdelavi domače naloge.** Domačo nalogo in pripadajoče programe sem izdelal sam.

## Priloge

**Programska koda.** Priložena je programska koda v datoteki **naloga1.py** . Slednja vsebuje kodo za preprocesiranje podatkov in implementacijo hierahičnega razvrščanja.

**Tekstovna datoteka dendograma.** Za boljšo preglednost.