**Total number of points to achieve is 60 (these points will be transformed into a final grade 0-20).**

**Part 1. PERL**

1. Find the mistakes in the Perl code. (5 points)

#!usr/bin/perl

use strict;

my @names = (“Jenny”, “Lisa”, “John”);

foreach my $one\_name(@names){

my\_subroutine ($one\_name);

}

sub my\_subroutine {

print "@\_\n";

}

2. What is the output of this perl script? (2 points)

#!/usr/bin/perl

use strict;

my $i = 0;

while($i >= 0){

$i++;

print $i . “\n”;

}

Answer: i will be printed from 1 till the finite integer number in perl

(Error due to the compilation and syntax mistakes)

3. What is the output of this perl script when parsing a FASTA formatted file and executing it the following way? (4 points)

perl myscript.pl > myfile.txt

#!/usr/bin/perl

use strict;

open FH, “file.fa”;

while(my $line = <FH>){

if ($line !~ (/^>/){

print $line;

} else {

print STDERR $line;

}

}

Answer: prints the protein/DNA sequence into the file “myfile.txt” and header names are printed with standard error

4. Hashes

* Describe the structure of a hash (1 point)

It is an associative array with “key - value” structure

* Give example of a hash. Write the initial assignment and the retrieval of a value by a key. (2 points)

my %hash = (name => Uni, password => 123);

print $hash{‘name’};

5. Make regular expression matching: (4 points)

Assume that words are in scalars or single entries in a line.

* "rat", "rot”, "rut” but not “rt” or “rxt” (1 point)

/^r[aou]t$/

* “/Cat” but not “/Cat1” (1 point)

/^\/Cat$/ or /^\/\w+\D$/ or /^\/\w{3}$/

* numbers 1,2,9 or 18 but not the other numbers (1 point)

/^([129]|18)$/

* two word characters, followed by one or more spaces, followed by a single digit, followed by one or more spaces (1 point)

/\w\w\s+\d\s+/ or /\w{2}\s+\d\s+/

**Part 2. Databases**

1. What is HUGO and HGNC? (1 point)

HUGO is Human Genome Organization and HGNC is The HUGO Gene Nomenclature Committee is the only worldwide authority that assigns standardized nomenclature to human genes.

1. What is a text parser?  (1 point)

Parsing is the process of analyzing a string of symbols, either in flat-file or XML. Text parser is a set of computer commands or a program that reads a structured or un-structured data and extract the specific information requested by the parser.

1. Describe the content of the OMIM database? (1 point)

Online Mendelian Inheritance in Man (OMIM) is a curated and continuously updated database of human genes and genetic disorders and traits, with particular focus on the molecular relationship between genetic variation and phenotypic expression. It is thus considered to be a phenotypic companion to the Human Genome Project.

1. What is a web service? (1 point)

In the broadest sense a web service is any service available on the World Wide Web. More commonly a web service is any service which is based on web technologies, which is intended for use by computer programs rather than people.

1. List important web services technologies? (1 point)

Representational state transfer (REST)

Simple Object Access Protocol (SOAP)

Web Services Description Language (WSDL)

1. Who are providing such services for programmatic access to biological databases? (1 point)

NCBI

EMBL-EBI

KEGG

**Part 3. Shell, NGS**

1. Put the steps in processing exomes in family sequencing in the correct order. (8 points)
2. Read QC
3. Read mapping
4. Mark/remove duplicates
5. Indel realignment
6. Base score recalibration
7. Variant calling
8. Joint genotyping
9. Variant annotation
10. What is the difference between whole genome and exome sequencing? (1 points)

Answer:In exome sequencing only the protein coding part of the genome is sequenced.

Alternative answer: only a targeted part of the genome is sequenced by selection step on microarrays.

1. Give two bash commands that bring you back to your home directory. (1 point)

cd ~

cd $HOME

cd

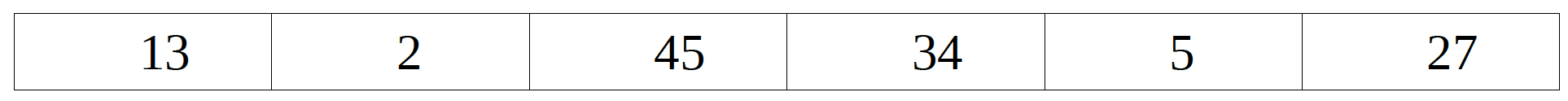
1. How to list all \*html files in the current directory? (1 point)

ls \*.html

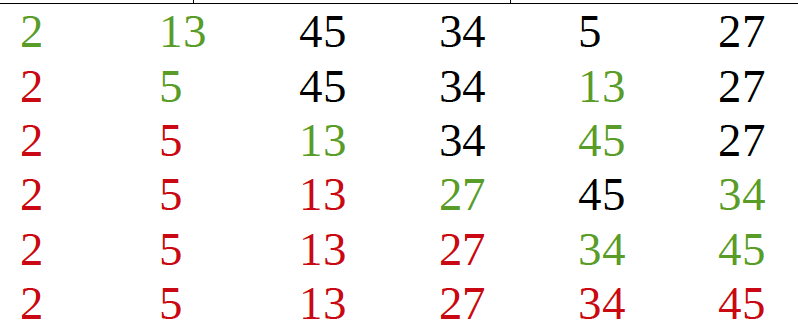
ls `pwd`/\*.html

**Part 4. Sorting, Searching and Alignment**

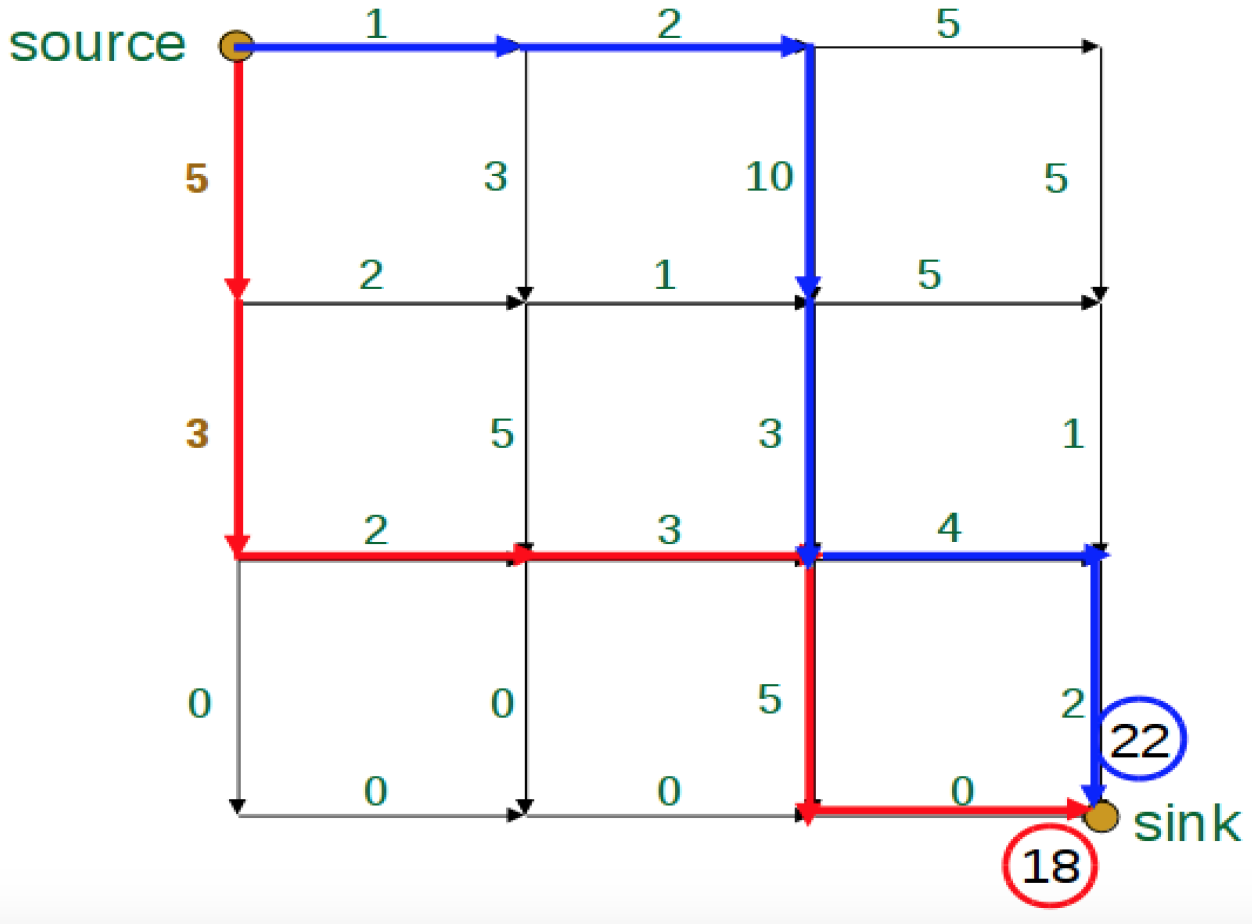
1. Perform a selection sort for the following (6 points)



Answer:



1. Given the Manhattan tourist Grid. Compute the greedy and optimal path. (6 points)



**Part 5. BWA**

1. Transform the given BWT string “WKRFWLOO!” into the Burrows-Wheeler-Matrix and write the original sequence. (10 points)

! W

FLOW!

KFLOW!

LOW!

ORKFLOW!

OW!

RKFLOW!

W!

WORKFLOW!

**Part 6. Computer architecture**

1. Describe the common architecture of a computer. Draw the connectivity of the components. (4 points)

CPU, Memory, Disc, Bus, Ports

1. Why can’t frequency of CPUs be increased any further? (1 point)

To much power consumption, the heat would destroy the CPU

1. What impact has the multi-core CPU architecture for the programmer? (1 point)

Parallel computation