1. (Recursion and Induction)

(a) \text{take} :: \text{Int} \rightarrow \text{[a]} \rightarrow \text{[a]}
\begin{align*}
\text{take} 0 \ \text{xs} & = [] \\
\text{take} \ n \ [] & = [] \\
\text{take} \ n \ (x:xs) & = x : \text{take} \ (n-1) \ xs
\end{align*}

(b) \text{drop} :: \text{Int} \rightarrow \text{[a]} \rightarrow \text{[a]}
\begin{align*}
\text{drop} 0 \ \text{xs} & = \text{xs} \\
\text{drop} \ n \ [] & = [] \\
\text{drop} \ n \ (x:xs) & = \text{drop} \ (n-1) \ xs
\end{align*}

(c) \text{suml} \ \text{xs} = \text{foldr} \ (+) \ 0 \ \text{xs} \quad \text{suml} \text{ computes the sum of a list}
\text{rowsum} \ \text{xss} = \text{map} \ \text{suml} \ \text{xss} \quad \text{compute row sum of the matrix}
\text{matsum} \ \text{xss} = \text{suml} \ (\text{rowsum} \ \text{xss}) \quad \text{compute sum of the whole matrix}

\text{listsum} \ [] \ [] = [] \quad \text{listsum sums two lists element by element}
\text{listsum} \ (x:xs) \ (y:ys) = (x+y): (\text{listsum} \ xss \ yss)

\text{colsum} \ [xs] = xs \quad \text{compute column sum of the matrix}
\text{colsum} \ (xs: xss) = \text{listsum} \ xs \ (\text{colsum} \ xss)

(d) \text{rank} 1 \ [x] = x
\begin{align*}
\text{rank} \ i \ (x:xs) & \\
| i <= n & = \text{rank} \ i \ lh \\
| i == n+1 & = x \\
| i > n+1 & = \text{rank} \ (i-n-1) \ rh \\
\text{where} \\
\text{lh, rh} & = \text{partition} \ x \ xxs \\
n & = \text{length} \ lh
\end{align*}

(e) \text{ac} :: \text{[String]} \rightarrow \text{[String]}
\begin{align*}
\text{ac} [] & = [] \\
\text{ac} xss & = \text{ac} [] \ xss \\
\text{acc} \ ys [] & = [] \\
\text{acc} \ ys \ (xs:xss) & = (zs): (\text{acc} \ zs \ xss) \\
\text{where} \ zs & = \text{ys} ++ \ xs
\end{align*}
(f) -- (dist xs ys) is True iff Hamming distance between xs and ys is 1.
    dist [] [] = False
    dist (x:xs) (y:ys)
        | x == y = dist xs ys
        | x /= y = xs == ys

-- hamming xss yss is True iff
-- the hamming distance between each pair of
-- corresponding elements of xss and yss is 1.
hamming [] [] = True
hamming (xs:xss) (ys:yss) = (dist xs ys) && (hamming xss yss)

-- adj xss is True iff all adjacent pairs of strings in xss
-- have Hamming distance of 1.
adj :: [String] -> Bool
adj xss = hamming xss (right_rotate xss)

2. (String Searching; 20 points)

(a) We have to show that \( c^j(Z_n 0) = Z_{n-j} 0 \). Proof is by induction on \( j \).

\[ j = 1: \] We have to show that \( c(Z_n 0) = Z_{n-1} 0 \), for \( n \geq 1 \). This follows from: (1) \( Z_{n-1} 0 \) is both a prefix and a suffix of \( Z_n 0 \), and (2) the only longer proper prefix of \( Z_n 0 \) is \( Z_n \), which is not a suffix.

\[ j + 1: \] Assume \( c^j(Z_n 0) = Z_{n-j} 0 \), where \( j < n \). We show that \( c^{j+1}(Z_n 0) = Z_{n-(j+1)} 0 \).
\[
\begin{align*}
c^{j+1}(Z_n 0) & = \{ \text{definition of } c^{j+1} \} \\
& = c(c^j(Z_n 0)) \\
& = \{ \text{induction hypothesis} \} \\
& = c(Z_{n-j} 0) \\
& = \{ \text{from the first proof} \} \\
& = Z_{n-(j+1)} 0
\end{align*}
\]

(b) Suppose \( p[0..k] \) is the core. From the definition of core,
\[
p[0..k] = p[12 - k..12]. \text{ Hence,} \\
p[i] = p[12 - k + i]
\]
Letting \( i = 3 \) and \( 12 - k + i = 5 \), we get \( k = 10 \). That is, if \( k = 10 \), \( p[3] = p[5] \). Since \( p[3] \neq p[5] \), \( k \neq 10 \). Similarly, using the fact that \( p[3] \neq p[9] \), we get \( k \neq 6 \). And, from \( p[5] \neq p[9] \), we get \( k \neq 8 \). Thus, the length of the core is not 6, 8 or 10.

(c) Choose \( q \) and make one pass over the genome sequence to compute \( val(p) = p \mod q \) for all substrings \( p \) of length 20 and less. Enter each \( val(p) \) in a hash table along with a pointer to the part of the genome sequence where \( p \) is a substring. To match a pattern \( r \) of
length less than or equal to 20, compute $val(r)$ and look it up in
the hash table. For all occurrences of this value, match the pattern
against the corresponding substring in the genome sequence.

For a pattern of length more than 20, let $r$ be its prefix of length 20.
Follow the same steps, as above.

3. (Relational Algebra; 10 points)

(a) The names of theatres which are showing PG movies in which Will
Smith is acting, is given by

$$
\pi_{\text{Theatre}}(\sigma_{p \land q}(R \bowtie T))
$$

where

$p$ is $\text{Actor} = \text{Will Smith}$
$q$ is $\text{Rating} = \text{PG}$

(b) Consider relations $R$ and $S$ in Table 1. Each relation has two at-
tributes, and just one tuple.

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Actor</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>Men in Black</td>
<td>Will Smith</td>
</tr>
<tr>
<td>S</td>
<td>Men in Black</td>
<td>Tommy Lee Jones</td>
</tr>
</tbody>
</table>

Table 1: Two relations

Take attribute $a$ to be Title. Now, $R \cap S$ is empty, so $\pi_a(R \cap S)$ is
also empty. However, $\pi_a(R)$ and $\pi_a(S)$ both have a single row (and
single column) with the entry “Men in Black”. So, $\pi_a(R) \cap \pi_a(S)$
has one row.